

Fraud, libel and the literature

A much-publicized attempt to show that the scientific literature is flawed is not itself above reproach. What science needs is less, not more, solemn attitudes to professional publications.

THE remarkable document on pp 207–214 of this issue should be read only in conjunction with the comment thereon by Dr Eugene Braunwald (pp 215–6) and the following explanation of *Nature's* role and interest in this unusual affair. Drs Walter Stewart and Ned Feder are two tenured members of the staff of the US National Institutes of Health (NIH) who have carried out an analysis of the research reports and other documents published in the scientific literature by John C. Darsee, who was discovered in May 1981, while a postdoctoral fellow at the Harvard Medical School, to have been guilty of deliberately fabricating data. So much was quickly established by committees of investigation set up by Harvard, by Emory University (where Darsee had previously worked) and by NIH. Stewart and Feder have attempted, by the scrutiny of Darsee's prolific published works, to identify errors, inconsistencies and other flaws that are unconnected with the known falsification of data. The essence of their conclusion is that errors, necessarily subjective in definition, are abundant; this, they say, reflects on the care with which Darsee's coauthors discharged their responsibilities, as well as on the management of scientific journals by their editors and referees.

As will be seen from Braunwald's comment, the methods and the conclusions of this unusual study are disputable; no doubt there will be other responses of the same kind to some of the other illustrations in Stewart and Feder's short list of examples. One obvious difficulty is that, by sticking woodenly to the decision that they would base their study only on the published literature, not asking those responsible for explanations, Stewart and Feder have no doubt been led to count as errors some occasions when authors have merely failed to make their intentions clear.

Yet the document will inevitably raise in readers' minds important questions about the reliability of the scientific literature and of the role of the process of publication in the practice of science. Is the literature in general as flawed as Stewart and Feder claim their "sample of convenience" to be? And, if so, why, and what can be done about it? Many will say that the first assertion has not been proved, so that the second question does not arise, but that is too complacent an opinion. There are also important questions about the pressure to publish that need attention, independently of the analytical part of what Stewart and Feder have written. So why not consider those issues in their own right, without publishing a document whose accuracy is disputed?

Notoriety

The truth is that the document by Stewart and Feder has by now acquired a notoriety comparable with the Darsee affair itself. An early version was submitted for publication in *Nature* in 1983; two reviewers (outside the United States) took the view that the manuscript, while of absorbing interest, would require that certain editorial questions of principle should be decided before publication could be seriously considered. The paper was accordingly sent for comment to a dozen or so of the coauthors identifiable even from the anonymous list of Darsee's publications in the accompanying bibliography, three of whom replied

through their attorneys with the opinion that publication would be defamatory, an opinion already given by *Nature's* own legal advisers.

Litigious readers may note, at this stage, that, at least under British libel law, it could well be libellous to publish an innocuous version of a manuscript accompanied by the intelligence that earlier versions had been defamatory of still identifiable people. Readers should therefore know that what follows has been discussed with some of those concerned, who have agreed that it should be published so as to bring out into the open an issue of which Stewart and Feder have recently made much.

Replication

Thanks to the marvellous powers of word-processors, the first version of the manuscript was quickly followed by many others. At a meeting in September 1984 with two of the co-authors and their attorneys, the editor of *Nature* was left in no doubt that publication of any of the versions then extant would be considered defamatory, an opinion with which he agreed. At the same meeting, a number of constructive suggestions were made for modifications that might make a publishable manuscript. In the event, the need to transact normal business, and the fear-some volume of comments and counter-comments that had accumulated, meant that it was not until the near the end of 1984 that an edited version of the manuscript which *Nature* considered to be acceptable was being transmitted by telefacsimile from London to the authors at NIH. When only 26 of more than twice as many pages had made the transit, the recipient authors telephoned hotly to say that the changes made the manuscript unacceptably anodyne, and that it would be withdrawn; they were told that *Nature* would be prepared to reconsider a new version of the original amended along the lines suggested if they failed in their expressed intention of having the original published elsewhere. The version now published is derived, mostly by abbreviation, from one submitted a year ago.

These tedious details are relevant to present circumstances because of the way in which Stewart and Feder have injudiciously, during the past year, sought to argue that the threat of libel suits has been used to suppress an important comment on the integrity of the scientific profession. This was the essence of their case to a congressional subcommittee in the United States last year, in a series of statements to the *New York Times* (26 April 1986) and, more recently and more damaging, in a letter (accompanied by an earlier unexpurgated version of the manuscript) asserting that publication had been prevented by the threat of libel suits; that letter was sent to all 1,800 members of the US National Academy of Sciences as well as to many of the authors' correspondents on quite unrelated matters. At a guess, upwards of 2,000 people have been circularized in this way.

Integrity

The issue thus raised has an important bearing on the conduct of the scientific literature, which depends for its integrity on its accessibility and openness; in principle, all scientists are free to offer their contributions. Dissent from accepted theories is allowed and even welcomed. Novel ideas may be published for the



sake of the stimulus they provide, but with no implied guarantee that they are in some absolute sense 'correct'. Moreover, authors are frequently made painfully aware by their experience that publications are hostages to fortune; it frequently occurs that the publication of a set of data and of a plausible hypothesis to account for them will stimulate others to produce further data whose effect is to contradict the hypothesis, to the chagrin of the original authors. Often, publications are found to contain self-vitiating errors, whose authors then have even redder faces. Such honest mistakes are an inescapable part of science.

So why does not this process of cut and thrust, personally hurtful to those concerned, require that scientific journals employ armies of lawyers vetting everything they publish? Because of the principle that, while it is a proper and necessary function of authors to correct what appear to be erroneous data or inferences in the literature in publications of their own, most keep to the convention (which is otherwise enforced by editors) that the identification of an error is not followed by the assertion that the perpetrator is a rogue, let alone by speculation about the motives that have engendered this condition. Stewart and Feder have complained that they could not publish what they chose because of "threats of libel". They have not understood that the unfettered right to publish scientific data does not equate with a right to denigrate others' characters.

For all this, what Stewart and Feder have written deserves close attention. At the very least, they have shown that a group of published papers do not provide a full explanation of the data recorded in them. Even if the "bizarre pedigree" they have found (see p.208) would be less puzzling if the ages given in the original paper were intended to mean something other than chronological age, there is nothing in the paper to suggest that simple readers would differ from Stewart and Feder in their interpretation of the paper. Similarly, while Braunwald's defence of the use of "historical controls" is entirely proper, it remains a fact that the procedure that should have been followed (Darsee cut corners instead) is not adequately explained.

Imprecision

How can imprecisions of this kind have found their way into published papers? Stewart and Feder are unforgiving in their opinion that Darsee's coauthors should have spotted them, and have got rid of them, but that asks a lot of flesh and blood. If the documents concerned were drafted in the first place by Darsee, often with the intention of rendering plausible a set of data derived from thin air by fabrication, it requires very little imagination to appreciate the difficulty of rooting out minor errors at source; those who spend time on committees know how much initiative rests with the one who writes the minutes. Stewart and Feder themselves appear to have spent a long time, assisted by the hindsight that Darsee was a fraud, scrutinizing just over a hundred publications. Braunwald also makes a cogent point in saying that the detection of the major error, the fraud, would have required the repetition of the experiments, defeating the objective of research by teams of people. But the general conclusion at which Stewart and Feder arrive, that coauthors should be more zealous in their scrutiny of what is about to be published in their names, anodyne though it may be, cannot be a bad prescription.

Coauthorship as such is a necessary consequence of the way that science is now practised, enabling people with different skills to work together on a single project; it is also capable of being abused. Stewart and Feder define as "honorary" those coauthors who have not been directly involved in the research reported in published papers. This is a restrictive definition, especially where the members of a research team include senior and less senior people. Most people, probably including Stewart and Feder, would accept the Braunwald position that a person whose role has been to conceive of a research programme, to raise the funds with which it is supported and to supervise progress, is legitimately a coauthor of the research. This is the

standard relationship between academic researchers and their graduate students, for example.

On more general grounds, Stewart and Feder are right to insist that the literature is corrupted if its users cannot be assured that all the authors of a publication have played some essential part in its development. There is evidence, mostly anecdotal, quite unconnected with the Darsee case, to show that what they call "honorary coauthorship" is commonplace. In molecular genetics, for example, those who supply clones of some genetic material for use by others are often listed among the authors of the ensuing research. Some coauthorship is political in flavour, as when the prime movers in international projects cite sleeping partners as authors for the sake of continued opportunities to investigate otherwise inaccessible phenomena. The practice that the head of a laboratory should as a matter of course appear as an author of all its publications, still accepted as a convention in some Indian laboratories, for example, is rapidly on the way out and should be helped on its way by Stewart and Feder's article, which will remind those concerned of the risks they run. Some journals, responding to the *samizdat* publication of that article, have gone so far as to decide that future coauthors will be required to declare their full involvement in work sent for publication (a practice, interestingly, followed as a matter of course by groups of authors from the Soviet Union). But by itself, such devices will not end the abuse of coauthorship.

Publications

How common are errors of the kind that Stewart and Feder categorize? And how serious are they? Their own study, some of whose conclusions are disputed, concerns an exceptional sample of the literature. But, shocking though it may seem, the experience of those concerned with the management of the literature is that errors of all the kinds listed are far from being rare. Even when authors do their best, honest mistakes crop up repeatedly. (Scientists are too often indifferent to the plain truth that awkward or ambiguous phraseology is one of the most common reasons why scientific publications are misunderstood.) Luckily, as the world knows, most of these difficulties are cleared up by the informal correspondence between interested parties which is now as much a part of the process of science as the formal literature. More serious faults are less easily dealt with. Instalment publication is often used to magnify a laboratory's output and, as all journal editors know, there are also many occasions when authors are less than diligent in referring to the intellectual sources of their work. Ignorance is often part of the explanation, but ignorance can cloak malign intent.

None of this bears directly on fraud, or on the prevention of it, but there are good reasons for believing that the ordinary defects of the literature arise from the undue importance now attached to the process of publication, largely because of the way in which the careers of younger scientists and the continued success even of successful laboratories have come to depend on what is sometimes known as 'output', which is another name for the number of research articles published, perhaps weighted by length and by the status of the journals in which they have appeared. Some assessment of this kind is the basis of what is called peer-review in the United States (strictly, not to be confused with the process by which journals base their decisions on what to publish on the opinions of external referees). As the economists would say, the research community has externalized the problem of making judgements about the qualities of its own members by brooding on bibliographies. Everybody's interest is to find a less slovenly way of making intellectual judgements about people and institutions. The recipe implicit in the Stewart and Feder argument, that zealous suspicion of everybody within sight is the way to ensure the integrity of the scientific literature, is of course a recipe for disaster, a road to general mistrust, a licence for every would-be whistle-blower and a means by which the literature would be made yet more solemn. Decoupling the literature from promotion prospects would be a better goal. □

AIDS campaign gets off to a sticky start

- Controversial advice to blood donors
- British troops advised to take AIDS tests

London

As part of an unprecedented £20 million health education campaign, the 20 million homes in the United Kingdom will receive a leaflet through the post, within the next three weeks, outlining the dangers of AIDS (acquired immune deficiency syndrome) and how to avoid its spread.

At the same time, there will be a stark television campaign conducted on both the commercial channels and those of the publicly funded networks of the British Broadcasting Corporation, drawing attention to the leaflets and their message "Don't die of ignorance". A special commercial has been made for cinema. The campaign illustrates the growing political pressure on the government from a medical profession that has become increasingly concerned about widespread ignorance of the disease and its frustration in being unable to find a cure.

That frustration was to provide acute embarrassment two days before the launch of the national education campaign. A statement by the British Medical Association (BMA) that those who had had more than one sexual partner in the past four years should refrain from donating blood — a voluntary act without financial reward in this country — caused much confusion and provoked hostility, principally from those trying to build up the resources of the national blood bank.

In response, the health officials, through the Department of Health and Social Security, were forced to play down the danger from AIDS on the eve of a campaign designed to do the opposite. A statement by the government's chief medical officer, Sir Donald Acheson, issued in the wake of the blood transfusion row, illustrated the dilemma.

The original statement read: "In only four of more than three million donations have antibodies to HIV [human immunodeficiency virus] been found in persons who deny being in any of the high-risk groups. This is striking evidence that

blood donors who are not in the high risk groups are very rarely infected with HIV." For that reason, men and women who were not in the main risk groups, even those who had more than one sexual partner in recent years, should continue volunteering to give blood, the statement continued. After reflection on this, the BMA withdrew its advice.

The emphasis of the educational campaign was different. Whatever we do, claims Norman Fowler, the Secretary of State for Social Services, it is likely that as many as 4,000 people in the United Kingdom will have died of AIDS in the United Kingdom by the end of the decade. Mr Fowler sees the object of the education campaign as stopping the spread of the

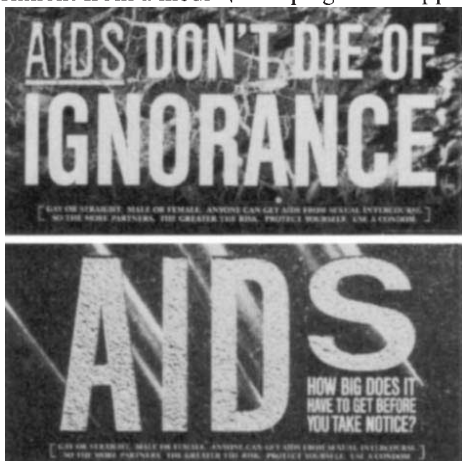
virus now and "for everyone to behave in a way which minimizes the chances of their being infected".

Despite the apparent contradictions in the two messages, the British medical authorities are seriously worried. The number of AIDS victims is to be published once a month and an interpretation of those statistics provided once a quarter.

One disturbing trend is the proportion of incidence of the disease that appear to be clustered around the London area. Seventy per cent of the 293 deaths in Britain up to the end of last year were reported in the London area. Panic press reports predicted that London would become an 'AIDS capital' while more reasoned articles suggested that the cases had clustered because of the medical facilities available in the capital — the first AIDS ward in Britain is to be opened this week at the Middlesex Hospital in London.

The dangers of the spread of AIDS in sexually active groups has focused attention on the armed forces. The 600 soldiers of the first battalion of the Queen's Own Highlanders have been urged to take tests for AIDS having just returned from an eight-week spell in Kenya, where the disease is thought to be prevalent amongst prostitutes.

Bill Johnstone



The government's poster campaign has been less cryptic than the advertisements puzzling television viewers.

Rifkin case thrown out

A US DISTRICT judge has thrown out of court an attempt by social activist Jeremy Rifkin to have declared illegal the US government's plans for regulating biotechnology. Rifkin has argued that the government's "Coordinated Framework" for regulation, published last June, was issued without adequate public participation, and that it lacked a required environmental impact statement. Judge Gerhard Gesell ruled Rifkin's complaint invalid on 22 December because the framework was a "far from successful effort to establish a common set of definitions" and because Rifkin's claim of injury rested on possible future government action.

Rifkin is undeterred by the apparent setback, arguing that the finding that the framework is not a formal "rulemaking" is a victory for his cause. It simply means, he says, that he must in future sue relevant agencies one by one, not *en masse*. □

DNA fingerprinting OK

DNA fingerprinting as a way to avoid or settle disputes over family relationships is given guarded approval by the UK Immigration Advisory Service in its annual report. The government-supported voluntary organization outlines the safeguards that it would like to see built into the application of the test to immigration disputes. Testing should not be compulsory, it says, but it would in most cases avoid the problems of trying to establish relationships by interviews. Blood samples would be collected overseas and sent for comparison with samples from relatives in Britain. □

Space centre waiting

BRITAIN has no cash to award new development space contracts because of the government's reaction to a 15-year space plan, completed at the end of last summer by the British National Space Centre (BNSC), is more than two months overdue. The delay is causing considerable embarrassment at BNSC, which was formed about 15 months ago to co-ordinate the British space effort. BNSC officials need approval soon if Britain is to participate in international projects such as the European space plane Hermes.

BNSC hopes to be able to negotiate its contributions to the European projects at an ESA meeting in the spring, when preliminary discussions will lay the foundations for a full ministerial meeting in June. France has already announced increased support for space, bringing its total expenditure on space to £700 million — nearly seven times that of Britain. □

Correction

THE Reagan administration's proposed budget for the fiscal year beginning next October is \$1.024 million million (*Nature* 325, 96; 1987). □

US courts and legislatures face implications of surrogacy

Washington

THE volatile issue of surrogate parents is once again bubbling. A test case in a New Jersey court will determine whether a contract requiring a surrogate mother to give up a child to the infant's genetic father and his wife is enforceable. Meanwhile, in New York state, a legislative committee is recommending drafting legislation setting up guidelines to regulate surrogacy, something no other state has yet done.

The New Jersey case arises from William and Elizabeth Stern's attempt to have a child through surrogacy because of Elizabeth Stern's medical condition, which makes child-bearing dangerous. Through the Infertility Center of New York, they contracted with Mary Beth Whitehead, a New Jersey housewife, for her to be artificially inseminated with William Stern's sperm. After being delivered of a baby girl, Mrs Whitehead initially gave her to the Sterns, but later decided she wanted to keep the baby. A court has given the Sterns temporary custody while deciding whether the contract between Mrs Whitehead and the Sterns can be enforced.

The New Jersey case embodies just one of the problems plaguing surrogacy. In the Stern-Whitehead case, both parties would like to keep the baby. But critics ask what would happen if the infant were deformed, when neither party might wish to keep it. Or what if the couple seeking the child simply changed their minds during the pregnancy?

US courts have already made some headway on the difficult issue of the rights of the genetic and the nurturing parents. There is a long-standing legal tradition that the woman who delivers the child is its mother. But H. Tristram Engelhardt, who deals with medical ethics at Baylor College of Medicine in Houston, emphasizes the distinction between a surrogate mother who provides the ovum and one who does not. Engelhardt says it is possible to consider the woman who 'lends' her uterus to nurture a fertilized egg as a temporary foster mother. According to Noel Keane, director of the Infertility Center of New York, a Michigan court has already recognized that the people providing the genetic material are a child's parents, regardless of who is actually delivered of the child.

But this genetic basis of determining parenthood has been challenged. Barbara Katz Rothman, a sociologist at Baruch College in New York, argues that the woman with the large belly and the swollen ankles is the mother, regardless of where the sperm and egg came from. She is concerned that, in the debate on surrogacy, the importance of a woman's nurtur-

ing role during pregnancy is being overlooked.

Critics of surrogacy are also concerned about the commercial aspects of the procedure. They regard the payment of a fee to a surrogate parent as tantamount to selling babies. Infertile couples have been willing to pay large sums in order to have children. Typically, a surrogate mother will receive a \$10,000 fee; medical and legal costs and payments to intermediaries can put the total cost as high as \$30,000.

Angela Holder, professor of clinical paediatrics and law at Yale School of Medicine in New Haven believes surrogate arrangements worked out between close family members or friends, where no money changes hands, are justifiable.

The proposals issued earlier this month in New York, by the state legislature's Senate Judiciary Committee, would give legal sanction to surrogacy and also regulate costs. The committee's report stresses the importance of the informed consent of all parties through independent legal representation. The report urges that courts review all fees involved, using a standard of 'just and reasonable compensation' and that, after successful conception from artificial insemination with the intended father's semen, the surrogacy contract should be binding. The child would be deemed the legitimate natural child of the biological father and his wife. Legislation containing these recommendations will probably be introduced this month.

New York is not the first state to consider laws on surrogacy, but although five states have considered legislation, no laws have yet been adopted. In any case, legislation is unlikely to end the controversy. Lisa Newton, a philosopher at Fairfield University, believes legislation will simply create more difficulties. She believes that legal principles to deal with surrogacy do not exist, and that thoughtful legal decisions to provide precedents for future cases are now needed.

Many countries are wrestling with surrogacy issues. In Britain, the Warnock Report in 1985 urged that surrogacy for profit should be banned, and there is now legislation to that effect. In West Germany, a government commission is expected this week to set guidelines for surrogacy; last September, a conference of lawyers decided that surrogacy did not violate human dignity if the gestational mother provided her own ovum. In France, a presidential panel has been holding debates on the issue, and the Council of Europe has been hearing arguments on a set of provisional principles concerning human artificial procreation. **Joseph Palca**

Official: UK's industry in a deep decline

London

A MAJOR political row erupted last week with the publication of the latest employment statistics which showed that Britain's manufacturing industry is in deep decline and is continuing to shed jobs. The statistics, prepared by the Department of Employment and based on its 1984 Census of Employment, showed that the job losses were largely concentrated in the traditional manufacturing areas, mostly outside the affluent South-East.

Two million manufacturing jobs are estimated to have been lost since 1979, with Scotland, Wales, Yorkshire, Humberside and the West Midlands bearing most of the burden.

The political criticisms were heaped on the government in the wake of the publication of the statistics that were hailed as proof that Britain has not properly invested in research and development and is increasingly becoming an assembly shop for foreign manufactured goods. Certainly the Cabinet Office, which is currently attempting to devise a new policy for British science, tailored by an influential secretariat, is highly critical of the amount of money available for development. The



research monies are sufficient it claims, but that available to exploit science is inadequate.

The conclusions of the Department of Employment from the statistics indicate that there will be little relief in the manufacturing sector in the immediate future. The latest figures on employment in manufacturing, says the department, show a decrease of 10,000 per month in the three months ending October. Although estimates have fluctuated from month to month, the average rate of decrease so far this year has been 14,000 per month. These revised estimates, the officials claim, "show a greater decrease in manufacturing employment than was previously thought". **Bill Johnstone**

Abortion-inducing drug alarms the right-to-life lobby

Washington

A NEW steroidal compound that may be used for treating cancer, glaucoma and even AIDS (acquired immune deficiency syndrome) is nonetheless at the centre of a political storm because of its most promising property: the ability to terminate early pregnancy with a single, 600-mg dose. The drug, produced by the French pharmaceutical company Roussel-UCLAF, is most commonly referred to by its original code name, RU 486. But opponents of the testing of RU 486 have called it the 'death pill'; they are asking that it should never be made available in the United States.

RU 486 is an antiprogesterone that is thought to compete with progesterone at the receptor site. The antihormone-receptor complex cannot trigger the actions normally provoked by the hormone. During pregnancy, the antiprogesterone increases contractility of the uterine muscles, an effect enhanced by prostaglandins. Administration of RU 486 will induce menses in woman who are not pregnant, and abortions — in most cases — in women who are.

RU 486 is also an antiglucocorticosteroid, which makes it useful in treating certain forms of Cushing's syndrome, a disease characterized by increased adrenocortical activity. Similarly, RU 486 may be used to treat other conditions related to excess corticosteroid production, such as lymphomas and glaucoma. By blocking cortisol activity at the receptor level, RU 486 can also interrupt the normal negative feedback control mechanism, stimulating the hypothalamic-pituitary-adrenal axis. This activation can enhance the effect of anticancer drugs, and alter the immune response, potentially relevant to treating AIDS patients.

Approximately 1,500 women around the world have now been given RU 486 to terminate pregnancy. The World Health Organization is conducting a multicentre trial in eight countries (China, India, Hungary, Sweden, Singapore, Vietnam, Italy and Scotland) using a combination of RU 486 and prostaglandins in tandem to induce abortions. The Population Council has also organized studies of RU 486's effectiveness as an abortion inducer both in the United States and overseas. Roussel-UCLAF has conducted its own clinical trials in France. In one such trial, recently reported, RU 486 successfully terminated pregnancy in 85 of 100 women treated within 10 days of their first missed period.

Supporters of the right-to-life movement in the United States are alarmed by the implications of RU 486, which they fear will make abortions easier to arrange,

and at the same time diminish a woman's sense that she is taking the life of her unborn child. A group of 6 conservative senators wrote to Secretary of Health and Human Services Otis Bowen last February, asking him to confirm that federal funds were not being used to investigate RU 486 as a potential agent for abortion.

In April, Bowen responded with a list of all projects using RU 486 sponsored by the National Institutes of Health (NIH). NIH have been scrupulously careful about not using their facilities in any way that might be considered related to abortion. Federal law prohibits federal money from being spent on abortion research. Tests on human subjects are all carried out before implantation of the fertilized egg.

Etienne-Emile Baulieu of the Université Paris-Sud, who has developed RU

486, would prefer that the drug should be called a contragestive agent rather than an abortifacient, on the grounds that it interferes with the establishment of pregnancy. But he admits that the political antecedents of the drug may impede its introduction in the United States, where liability issues could also price the drug out of a potentially enormous market.

There are other than social problems for RU 486. The drug has a relatively long half-life — 20 hours — allowing it to remain in the body longer than would optimally be desired. Heavy bleeding follows treatment in some women, which means that taking the drug without medical supervision could prove dangerous.

Despite these problems, other companies are interested in the potential of RU 486. Schering AG in West Germany has developed a similar compound now in clinical trials. And Roussel-UCLAF plans to submit licence applications for RU 486 this summer in France and Sweden.

Joseph Palca

New Australian law on embryos still confuses researchers

Sydney

WHAT is an embryo? The answer could provide a solution to some of the issues raised by the Infertility Medical Procedures Act which came into effect last August in the state of Victoria, Australia. The threat of four-year gaol terms contained in the act brought *in vitro* fertilization (IVF) research in Victoria to a complete halt and prompted Dr Alan Trounson, the leader of Monash University's world-renowned IVF research team, to issue an ultimatum that he and his team would go overseas within six months if they were not allowed to continue research. Three months have now passed (see *Nature* 232, 748; 1986).

The legislation forbids the destruction of embryos but it never actually defines an embryo and at what point a fertilized egg becomes an embryo, or even what fertilization is. According to Professor Carl Wood, chairman of the department of obstetrics and gynaecology at Monash University and head clinician of the Monash/Queen Victoria Hospital IVF programme, all efforts towards a resolution of the researchers' crisis have been stalled by the notion that an embryo comes into being the moment the sperm head passes into the shell of an egg. Although this is an essential step, there are others that might more logically mark the beginning of an embryo's existence, such as 20 hours after fertilization when the male and female genetic material fuse together or 14 days which mark the first appearance of cells that will have a lineage to the fetus and eventually to the new human. Before the 14th day, all the cells that grow from the

fertilized egg go to make up the support systems for the fetus such as the fetal sac and the placenta.

The Standing Review and Advisory Committee on Infertility chaired by Professor Louis Waller has power to adjudicate on the interpretation of the legislation. A request from the Monash scientists for clarification of terms such as embryo and fertilization used in the legislation, as well as what constitutes an abnormal embryo and when such can be discarded, is now before the Waller committee. Adoption of a 20-hour limit for embryos should clear the way for approval of two proposed experiments before the Waller committee. This would be sufficient to ensure that the Monash team stays.

One experiment promises the development of a new method for the treatment of male infertility, fertilization by the micro-injection of single sperm, and the other is on the gamete freezing method which involves the freezing and thawing of egg cells. Both experiments involve the fertilization of eggs without implantation — an illegal act while fertilized eggs are classified as embryos.

The Monash team would like to see the adoption of the term 'pre-embryo' for the initial 14-day period as used in Britain; this might eventually lead to a broader range of permissible experimentation.

There is still no indication from the Waller Committee on when it will make its rulings. Trounson sees some irony in the situation. If such a law had been in force six years ago, thousands of families would still be childless.

Charles Morgan

USA, Australia and China collaborate over Pacific Ocean

San Francisco

A RESEARCH programme deemed to be one of the biggest and most sophisticated studies ever assembled to probe the dynamics of storm systems and their influence on upper atmosphere chemistry has just been launched.

The world's highest, coldest thunderstorms, which billow into the stratosphere every January and February above the tropical Arafura Sea north of Australia and the neighbouring Indonesian archipelago, may be the key to a vexing meteorological problem: how gases move between the upper and lower atmosphere.

A US National Aeronautics and Space Administration (NASA) ER-2 jet at an altitude of 20 km will test a novel hypothesis proposed about five years ago that these tropical storms in this restricted portion of the globe throw large volumes of lower atmosphere to higher levels and simultaneously preserve the extraordinarily low humidity of the lower stratosphere. The aircraft test is part of a US\$10 million NASA project called Stratosphere-Troposphere Exchange Project (STEP). If the theory is supported, the warm seas northeast of Australia may contain the prime meteorological channel through which the bulk of atmospheric contaminants, including industrial discharges such as ozone-destroying fluorocarbon products, reach the stratosphere. Reginald Newell, professor of meteorology at Massachusetts Institute of Technology and an adviser to the programme, has called the area a "stratospheric fountain".

The single-engined ER-2, a research version of the US Air Force TR-1 high-altitude reconnaissance jet derived in turn from the famed Lockheed U-2 spy planes, left the NASA Ames Research Center in California on 8 January on a flight to Darwin via Hawaii and Guam. Its 1400-kg package of 16 instruments in the fuselage and two under-wing pods is the heaviest science payload the aircraft has ever carried.

Waiting at an Australian Air Force Base near Darwin to begin a month-long assault on large tropical storm buildups is a multinational research team for STEP. The project, five years in the making, is coordinated with two other studies, the Australian Monsoon Experiment (AMEX) and the Equatorial Mesoscale Experiment (EMEX). Participating agencies include NASA, the US National Oceanic and Atmospheric Administration (NOAA), the Australian Bureau of Meteorology, and several university-based groups. In addition, the crew of a research vessel from the People's Republic of China will help, launching

radiosonde balloons and carrying a US-supplied weather radar to track storm systems for the combined effort.

While the sailplane-like ER-2 flies above, and into, the anvil-shaped tops of massive storms at 60,000 feet and above, other American and Australian aircraft will inspect lower altitudes. The measurements from the Chinese ship will augment a network of a dozen radiosonde and six radar sites along northern Australia.

NASA aims to sort out the precise nature of an atmospheric transport system known as the Walker Circulation, a convective cell with predominantly east-west orientation. STEP will examine the western end of the Walker Circulation as it gathers moist, warm low-level air flowing from near the coast of Mexico, lifting it up over the Western Pacific to feed dry, cold winds blowing back to the east.

Conventional wisdom has it that tropical thunderstorms flatten out at 50,000 feet or so. "We've seen them to 60,000 feet here, and radar suggests they go even higher," according to NASA meteorologist, Edwin Danielsen.

It is an issue only direct measurement can settle. The ER-2's pilots will fly above and into the cloud sheets swept outward of thunderstorm tops, giving them their characteristic anvil shapes as stormclouds overshoot points of equilibrium, flatten out, and extend downwind. The high-altitude cloud decks, rapidly losing heat to the dark sky above, and absorbing radiant heat from the sea, should become vigorously convective. As the NASA team sees it, they will become heat engines, pumping tropospheric air up, while condensing water into ice that falls back down.

Ozone, photochemical nitrogen compounds, and cosmogenic isotopes such as beryllium-7 and phosphorus-32 characteristic of the upper stratosphere, will be compared to aerosols, water vapour and water ice, carbon monoxide and radon daughters swept up in the tropopause. Temperature, wind speeds and pressure will all be carefully charted.

The project offers another big bonus. Pending approval of Chilean authorities to use an airfield at the southern tip of South America, NASA and NOAA officials plan to send the plane over Antarctica in August and September. This is when the height of the recently reported annual south polar ozone depletion is expected to occur (*Nature* **300**, 686; 1982). The Antarctic mission may uncover the role of man-made pollutants, delivered to the upper atmosphere through such mechanisms as the western Pacific stratospheric fountain, in piercing the hole in the ozone.

Charles Petit

Italy's nuclear forum is postponed

London

A CONFERENCE billed to "decide the future" of nuclear power in Italy has been postponed, throwing the Italian nuclear lobby into confusion.

Prime Minister Bettino Craxi's Socialist-Christian Democrat coalition government has equivocated on nuclear energy in public, but privately is said to back the country's recent "energy plan" to build two or three pressurized-water reactors (PWRs, Italy's first). The conference plan was supposed to be the political solution. To help satisfy public opinion, the government was (and in principle still is) to hold an open, three-day parliamentary conference covering all aspects of nuclear power, particularly safety. This was due to take place next week; but it always seemed to be a dangerous balancing act destined to receive a great deal of media attention, but with no clear powers of decision, so the postponement is perhaps unsurprising.

The Chernobyl accident last year caused an uproar in Italy. Although the country was only slightly exposed to fallout from the explosion of the Ukrainian reactor, Italian politicians (who may soon face an election, were careful to listen to public complaints, and one after another of Italy's many political parties pledged itself against nuclear power. By late last year there was a political lobby large enough to pose a serious threat to the country's fledgling nuclear power programme.

Not only are the plans uncertain for new PWRs in Italy, but also for existing reactors and those under construction. Only four per cent of Italian electric power is nuclear-generated. (Italy is dependent upon foreign imports for some 80 per cent of its energy, making it one of the highest import dependences in Europe.) Two boiling-water reactors are in an advanced state of construction at Montalto di Castro (between Rome and Pisa), and work has begun on three PWRs sited east of Turin in the Piedmont, east of Milan in Lombardy, and south of Rome. Allowing for the decommissioning of older reactors, this would take Italy to around 25,000 million kWh per year of nuclear generating capacity by 1995, which would still amount to only 10 per cent of the country's requirements for electricity by that date, according to the Italian nuclear and alternative energy agency ENEA. The strategy, according to the president of ENEA, Professor Umberto Colombo, is to "hold the position" until renewable and fusion energy becomes available, although, he

admits, that could be "forty to fifty years away".

Safety will be the main issue at the conference, but the Italian commitment to the fast breeder reactor will also be in question. Italy owns a one-third share in Europe's — and the free world's — largest such reactor, the 1,300 MW Superphénix (at Creys-Malville on the French Rhône river), and is also party to a Europe-wide agreement on fast reactor research and development. But the Italian Communist Party (a party of Euro-communists who distance themselves from Moscow, unlike their French equivalents) has set itself firmly against fast reactors. The Communists are by far the strongest of Italy's left-wing groups, and Colombo is "sorry" that Italian withdrawal from fast-breeder research now seems "likely", although continued participation in Superphénix is ensured (for the electricity it provides Italy). The most immediate impact of withdrawal from research would be the end of the so-called PEC 120 MW (thermal power) experimental fast breeder, designed to study fuel safety, which is now 70 per cent complete on a site midway between Bologna and Florence.

Craxi's conference to "decide" on even this small amount of nuclear power and nuclear research has already begun in one

sense: the industry ministry in Rome has sent questionnaires to 150 environmental groups, nuclear power and other institutions, asking for their forecasts of future electricity demand and supply in Italy, their estimates of environmental damage and risk, and for proposals for an institutional framework for managing Italian energy and any accidents that might arise.

The conference proper will consist of analysis of the replies, reports by invited experts from Italy and the rest of the world and most critically a closing assessment by the government.

Then, an interparliamentary commission will attempt to reach conclusions, but as Colombo predicts, the short conference is likely to leave a very divided field. An assessment of likely votes is as likely to sway politicians as an assessment of technical arguments, and for the Italian nuclear power lobby a public "*bella figura*" at the conference must be essential.

Meanwhile, however, the spring date approaches when Craxi (a socialist) has promised to hand over the premiership to his Christian Democrat colleagues. He has refused to do this once before, and many think the same thing may happen again, this time precipitating a general election. That could mean Italy's energy future being decided in the white heat of Italian politics.

Robert Walgate

NASA mosquito watch threatened

Washington

PREDICTING malaria outbreaks from space may be possible using a model being developed by the National Aeronautics and Space Administration (NASA). The model will use remote-sensing images taken by Land Remote Sensing Satellites (Landsats) of malaria-prone regions to forecast when and where outbreaks are likely to occur. But the budgetary problems that have recently befallen the Earth Observation Satellite company (EOSAT), Landsat's owner, may mean that projects like the malaria watch will have trouble getting off the ground.

If money problems can be solved, the first phase in the development of the malaria model will take place this summer in the rice fields of Northern California. A local species of mosquito is closely related to the ones that spread malaria elsewhere, although California's control measures have rendered the state virtually free of the disease. According to project chief Paul Sebesta, researchers at the NASA Ames Research Center will be collecting "ground truth" data to work out the relationship between the satellite images of ground reflectance patterns and the variables that influence the proliferation of malaria-bearing mosquitoes. Such ground variables include surface water distribution, rainfall patterns, drainage, vegetation type and temperature. Once these relationships are established, it will be possible to predict both the extent of the mosquito population and how long it will thrive.

The second field trial of the new technique will begin in 1988 in the southwestern oceanic plain of Mexico, near Tapachula, a site with a large malarial mosquito population. The Mexican location was selected during a workshop between NASA scientists and world health officials held early last year at the Uniformed Services Medical Center in Bethesda, Maryland. Criteria for choosing a site were that it should have only one species of mosquito serving as the malaria vector, be politically stable and easily accessible. Costa Rica, Turkey, Pakistan and China were also considered, and they are likely to top the list for future trials. If successful, the model will be turned over to national and international malaria control authorities. Carol Ezzell

Landsat remote sensing project short of funds for satellites

Washington

As a result of a delay in 1987 budget negotiations between Congress and the National Oceanic and Atmospheric Administration (NOAA), work on the two newest Land Remote Sensing Satellites (Landsats 6 and 7) has ground to a halt.

The Earth Observation Satellite Company EOSAT, dependent upon a NOAA subsidy to construct the new satellites, has been forced to terminate Landsat development subcontracts. These contracts were held by the Santa Barbara Research Center of Hughes Aircraft Company and General Electric's RCA Astro-Space Division, among others. Termination of the subcontracts will mean layoffs of up to approximately 200 scientists and technicians for each major contractor, and the loss of 17 marketing support personnel for EOSAT. Although the two currently operating Landsats 4 and 5 will not be immediately affected, a continued lack of funds could drive EOSAT into a worse predicament.

For the past five years, NOAA has been a favourite target of Reagan administration budget cuts, struggling each year to get funding. EOSAT has been prone to the vagaries of the forces influencing the NOAA budget since it was formed to

commercialize the Landsats. In 1985, when EOSAT took over the Landsats, Congress decided that EOSAT would receive \$250 million in start-up funds, to carry EOSAT through the deployment of Landsats 6 and 7. So far, each year EOSAT's subsidy has been deleted from the NOAA budget by the Office of Management and Budget (OMB), only to be put back in by Congress.

For 1987, EOSAT was scheduled to receive \$87 million, but OMB did not approve NOAA's budget. Congress then approved the appropriation of \$27.5 million of new money for EOSAT, but demanded that NOAA provide a detailed plan for the use of the funds. The request for a new plan was prompted by delays in the Shuttle programme since the Challenger accident left the Landsats with no launch vehicle.

The current, unapproved NOAA plan calls for a \$209.2 million total subsidy, down from the original \$250 million, for the development of only one satellite instead of two. Launch is proposed for 1989 using a Titan 2 missile. This means a budget in 1987 of \$62.5 million. EOSAT says it requires a minimum of \$69.5 million in 1987, and is still pushing for plans for two new satellites.

Congress, EOSAT, and NOAA are still trying to reach a compromise. Once an agreement has been settled, however, EOSAT will have to renegotiate the contracts with its subcontractors, a task that could set the project for the new Landsats back by nearly two months. Meanwhile, EOSAT, Landsat and the fate of any projects dependent upon Landsat data are caught in budgetary limbo. Carol Ezzell

Japan overtakes Soviet Union in research spending league

Tokyo

JAPAN'S investment in research and development continues to soar, with industry providing the lion's share, although basic science is still short of funds. Research spending rose more than 12 per cent in the last fiscal year (ending March 1986) to surpass that of the Soviet Union; Japan now ranks second only to the United States, according to figures from the Management and Coordination Agency.

Private industry has dominated recent growth in investment while the share of funding for the universities and basic science continued to lag. But in the life sciences, and in particular genetic engineering, there were large increases in both the public and private sectors.

It is a characteristic feature of Japanese science and technology that research funding is provided largely by industry. As a result, the research is predominantly applied and product-oriented. Although some Western governments might be envious of such a situation, it causes great concern in Japanese government circles. The chronic national debt, which now stands at ¥152 million million (nearly \$1 million million), has prevented any significant increase in government funding for basic research, and the gap in funding between basic and applied research continues to widen.

Total spending on research and development, including outlays for the humanities and social sciences, reached ¥8.8 million million (\$58,000 million), 2.8 per cent of Japan's gross national product (GNP). Nearly 80 per cent of this was provided by the private sector, which increased its expenditures by 15 per cent over the previous year.

The big spenders in industry were the electronics, chemical and car industries, the electronics companies investing ¥1.8 million million. Steel and ceramic industries also increased research funding by more than 25 per cent.

In the energy sector, nuclear power research rose 10 per cent to ¥350,000 million. With electric power companies raking in profits as a result of low oil prices and the high yen, and with the government and industry holding ambitious plans for expansion of nuclear power, funding for this type of energy research scheme is bound to increase. Outlays for research on natural sources of energy (such as wind and

solar power) on the other hand remained stagnant or decreased.

Space and marine research, much of which is funded by the government, show healthy increases. But research spending in the universities, where much of Japan's basic research is carried out, did not even

Spending on research and development

Country	R&D expenditures (¥million million)	% of GNP
Japan (1985)	8.89	2.77
United States (1985)	25.95	2.72
Britain* (1983)	2.37	2.18
West Germany (1985)	4.23	2.84
France (1984)	2.58	2.10 (1982)
Soviet Union (1984)	7.69	3.73 (1982)

*The figures for Britain do not include the humanities and social sciences. The equivalent figures for Japan are 8.12 million million yen and 2.53 per cent of GNP.

keep up with inflation. Funding for science (including engineering and agriculture) rose a mere 1.1 per cent to ¥1 million million. With more than 120 scientists and engineers in the universities, more than 80 per cent of these funds were swallowed up by salaries alone.

Basic science did receive a boost, but not from the government. Industry has begun to perceive the need for basic research (or at least for the kind that might

lead to applications), and privately-funded laboratories for basic research have been sprouting up all over the country (*Nature* 311, 404; 1984). Industry's contribution to basic science increased by more than 20 per cent in fiscal year 1985 to ¥350,000 million, a large proportion of this coming from the chemical and pharmaceutical industries. Nevertheless, the proportion of Japan's total research and development funds allotted to basic science continues to decline. The picture in the universities is not all gloom, though, and funding for the life sciences has increased by leaps and bounds.

Prime Minister Yasuhiro Nakasone's enthusiasm for the life sciences and basic research in this field is destined to give these disciplines a further boost this year if his "Human Frontiers Science Program" gets off the ground. Promoted as Japan's answer to the US Strategic Defense Initiative and Europe's Eureka, the frontiers programme is expected to support basic research into the functions of living organisms (including man), such as mechanisms of the brain, and the nervous and immune systems. But even here Japan has its eye on future applications, such as intelligent robots and artificial muscle.

But it is not clear how the funds for such a major initiative will be raised, although tax reform and privatization of the national railways may help to boost government finances in the long term.

David Swinbanks

Cold spell taking its toll in the Eastern bloc countries

London

THE current cold spell has caused considerable disruption to the Soviet Union's already over-stretched energy supplies. With temperatures the lowest of the post-Stalin era, there has been inevitably an increased demand for electricity. But, even before the cold set in the situation was already serious, with rationing and staggered working hours in force in many of the union republics (see *Nature* 324, 403; 1986). The Chernobyl disaster, and the shortfall of hydroelectric capacity due to the drought last summer, are being cited as the main causes. But other factors are clearly involved. A politburo meeting last August noted the need to step up the refitting and overhauling of existing power stations, the stock-piling of fuel at power stations, and, in particular, the introduction of fuel and energy-saving measures, including the "prewinter preparation of dwelling houses".

The latter point is significant. Shoddy workmanship in the construction of apartment blocks (a major target of the new policy of openness) and infrequent and inadequate maintenance are responsible

for considerable waste of heat, while, according to a spokesman for Moscow city council, many of the new blocks were simply not designed for such cold weather.

But the extra fuel needed is not readily available. Coal trains are snowed in and gales in western Siberia have brought down the exposed power lines that serve the oilfields, halting the extraction of oil.

Elsewhere in Eastern Europe, the situation has been similar, with transport paralysed, massive demands on already over-stretched energy supplies. In Hungary, an inter-ministerial coordinating commission has been set up to deal with what is officially termed an "extraordinary situation", and rescue work is being undertaken jointly by Hungarian and Soviet troops. (Normally, except during Warsaw Pact exercises, Soviet army units in Hungary keep an extremely low profile.) Bulgaria, which several weeks ago had already urged stringent energy-saving including fines for excessive use, and Romania, where prolonged power cuts occur regularly every winter, expects considerable difficulties in the next few days.

Vera Rich

Danube water-economy plans stir up environmental opposition

London

LAST August, when the USSR Supreme Soviet and Central Committee of the Communist Party cancelled the 'project of the century' — the diversion southwards of the north-flowing rivers of Russia and Siberia — the decision gave impetus to the growing campaign for the cancellation of another, less-publicized water-management project, the Danube-Dnieper irrigation canal. The canal, together with auxiliary works including two hydroelectric stations on the lower Dnieper, is known as VHK, the Ukrainian acronym for 'water-economy complex'. But there are growing fears among Ukrainian scientists that VHK may prove a false economy.

Addressing the Soviet Academy of Sciences recently, the president of the Ukrainian academy, Academician Borys Paton, claimed that preliminary estimates put the construction costs at more than 30,000 million rubles and that the canal would also create "negative ecological consequences" including the cutting off of estuaries from the sea, the need to install expensive and sophisticated technology to purify the waters of the Danube, the flooding of vast tracts of agricultural land, and the problem of rendering harmless 5 or 6 km² (presumably annually) of highly mineralized drainage run-off. An expert commission of the USSR State Planning Commission has demanded a revision of the plans, which have not been formally ratified. But, Paton said, "most regretably, as happens in our country", the preliminary work has already started.

The main purpose of VHK is to develop the agricultural potential of southern Ukraine and Moldavia. Paton himself is highly critical of current Soviet investment policy in land improvement, which, he told a special session of the Ukrainian academy last year, spends a disproportionate amount on irrigation, to the detriment of anti-erosion and anti-acidification work, desalination and the like.

The use of Danube water for irrigation, in the opinion of the anti-VHK lobby, could cause considerable damage both to agriculture and the river-estuary coastal ecosystem. Ironically, the Kiliya channel, the northernmost arm of the Danube delta, from which the proposed canal would presumably start, is now being developed as a nature reserve of unique ecological significance.

Although the Danube is an international waterway, there is at present no international body responsible for water quality. The Danube Commission, established in 1948 by the Danube Convention which the Ukrainian SSR signed as a sep-

arate state (the Moldavian SSR was represented by the Central Soviet government) is concerned only with problems of navigation, and thus only with pollution resulting from ships. There is, so far, no international body responsible for land-based pollution, although much Central European industry is situated along the Danube. In December 1985, a meeting of representatives of the riparian countries was convened in Bucharest to discuss a unified environmental strategy for the river, but this produced little more than a preliminary statement of intent. The opponents of VHK say that the scheme could ruin the "recreational natural resources" of the Ukrainian republic, in which there has been considerable investment in recent years, while, according to Professor B.P. Zahoruyko, rector of the Odessa Maritime Fleet Engineers' Institute, the project could cause "no little harm" to river transport and fisheries. Zahoruyko is particularly alarmed at the proposal to isolate the Dniester estuary and the Belgorod-Dnestrovskii harbour from the Black Sea by a dam.

Whether the protests of the Ukrainian scientists will be as effective as those of the Russians may eventually depend on what factors really decided the cancellation of the north-south diversion. According to Academician Aleksandr Yanshin, a vice-president of the Soviet Academy of Sciences, the decision was made on strict scientific criteria; suggestions that public concern over the environmental consequences played a significant part are "a long way from the truth". Less official conversations with Soviet academicians, including ecologists, suggest that protests from the representatives of the humanities within the academy and from the group of Russian writers known collectively as the 'villagers' may have helped to sharpen the issues for the experts who took the final decision.

In Ukraine, however, the writers have been less concerned with environmental issues, and so were soundly scolded during the Congress of Ukrainian Writers last summer by the first secretary of the board of the Ukrainian Writers' Union, Pavlo Zahrebelnyy. They should emulate the "boldness" of their Russian colleagues, he urged, and speak up against the "senseless planning" that is inflicting such harm on the environment. If the Ukrainian writers and representatives of the humanities fail to respond to Zahrebelnyy's call, then the ultimate decision on the Danube-Dnieper VHK may well indicate what part public pressure from concerned individuals really played in the cancellation of the north-south diversion.

Vera Rich

Biotechnology firm turning the corner?

London

CELLTECH, Britain's only major biotechnology company, together with its two associated companies, threatens to show its first profit in the coming financial year. Although the dominant patent position of Genentech has forced Celltech to stop its development of tumour necrosis factor as a human therapeutic product, the company is optimistic about gaining a share of the market for tissue plasminogen activator, as it believes that the patent position outside the United States is still open.

Reporting a turnover of £7.6 million with a loss of £0.7 million for the financial year that ended in September 1986, Celltech emphasized that it remains the world leader in bulk production of monoclonal antibodies, which account for most of the £5 million product sales of Celltech, Boots-Celltech and Apcel last year. Hopes run high for sales of a monoclonal-antibody-based test of ovulation. Measuring the ratio of two urinary metabolites of oestrogen and progesterone, the test is expected to become available in infertility clinics within months. But, in collaboration with London International, Boots-Celltech hopes to develop a dipstick-like version of the test to be sold over the counter as an anti-conception aid.

Celltech's chief executive, Gerard Fairtlough, placed particular emphasis last week on two relatively new projects. One is the development of the tissue inhibitor of metalloproteinases as a therapeutic substance. The inhibitor is part of the natural mechanism by which the turnover of connective tissues is controlled and is expected to be useful in the therapy of rheumatoid arthritis and other connective tissue disorders.

The other highlighted project combines Celltech's recombinant DNA and monoclonal antibody expertise. Financed by American Cynamid to the tune of £5 million over two years, it is aimed at cancer detection and treatment with genetically tailored antibodies to tumours, complexed to toxins, drugs or radionuclides.

In a report on Celltech released this month, stockbrokers Wood Mackenzie place the company within the top ten in the world in terms of revenues and describe it as a "robust competitor". Celltech is likely to develop very rapidly over the next three years, says the report, and should be in a position to offer its shares to the public "in 1987 or 1988"; Gerard Fairtlough concurs, adding that any decision to go public will depend on the state of the market and the needs of the company and its shareholders.

Peter Newmark

Gene manipulation in Britain

SIR—Joseph Palca's article "Living outside regulation" (*Nature* 324, 202; 1986) reports that a genetically manipulated yeast developed in the United States may be tested in England at least in part to avoid the US regulatory process. The article might be interpreted by some readers as implying that in England regulatory requirements for this type of work are more relaxed than in the United States.

Although I am not in a position to make direct comparison between regulatory requirements in the United States and Britain, I would like to make the point that in Britain the safety issues raised by genetic manipulation have traditionally been handled in a balanced and flexible way as exemplified by the framework for notification, risk assessment and laboratory containment established by the former Genetic Manipulation Advisory Group (GMAG) and the Health and Safety Executive (HSE) in the 1970s.

The Advisory Committee on Genetic Manipulation (ACGM), HSE and other involved government agencies intend to continue this tradition as genetic manipulation finds increasing application outside the laboratory.

Existing requirements in this area are for those proposing to carry out large-scale fermentation of recombinant DNA organisms to notify HSE in advance according to guidelines issued by GMAG in 1979 (revised 1982). Each proposal is then considered on a case-by-case basis by HSE and ACGM. Also, revised and more detailed guidelines on the risk assessment and notification of this type of work are under preparation. In common with genetic manipulation laboratories, large-scale sites are subject to inspection under the Health and Safety at Work etc. Act 1974 by HSE's specialist microbiology inspectors who are also available for advice.

For the environmental release of recombinant organisms, guidelines were issued in 1986 calling for prior notification and setting down risk assessment factors for such work. Notifications are considered by ACGM, its Planned Release Sub-Committee, HSE and other relevant government agencies (the Department of the Environment, the Ministry of Agriculture, Fisheries and Food (MAFF), the Department of Health and Social Security and the Nature Conservancy Council) in advance of any release.

Also the Health and Safety (Genetic Manipulation) Regulations 1978 are under review and this review is expected to take full account of the growing importance of large-scale fermentation and release into the environment. For genetically manipulated organisms as pesticides, the requirements of the Control of Pesticides Regulations 1986 are relevant. The

Food Act 1984 places a duty on the processor/manufacturer/retailer to ensure that food offered for sale for human consumption is safe. There is also a notification scheme for novel foods operated by MAFF in the United Kingdom that would apply to the use of free-living microorganisms. Data on foods under this scheme are evaluated by the Advisory Committee on Irradiated and Novel Foods.

It is important that the oversight of genetic manipulation by government strikes the right balance between avoiding undue restraint and ensuring that appropriate consideration is given to potential concerns. On the international front it is to be hoped that the recently published OECD report *Recombinant DNA Safety Considerations* will do much to harmonize national regulatory positions. Meanwhile, reports of regulatory difficulties in one country do not automatically mean that others have the balance wrong.

B. P. AGER
(Secretary)

*Advisory Committee on
Genetic Manipulation,
Health and Safety Executive,
Baynards House,
1 Chepstow Place,
London W2 4TF, UK*

An open question?

SIR—Michael Wagner (*Nature* 323, 664; 1986) implies that a sincerely held belief has the same status as an informed opinion and cites Feyerabend¹ in opposition to Scriven while neglecting to mention that Feyerabend is a 'philosopher' who therein contends that there are different paradigms of rationality and reliance on faith healers is justified because our own paradigm of scientific rationality is founded on irrational elements. Elsewhere², Feyerabend maintains that science without sensory experience is possible by means of telepathic communication with a somehow pre-existing, pre-programmed computer. Such is the present state of philosophy.

In countering Marks' claim that there are no theories of paranormal investigation Wagner asserts that "there are, of course, philosophically cautious theories of the paranormal and different models of objective transcendence". One reference given to justify this assertion accepts reincarnation as a fact at the outset and goes on to say that we interact with other realities via higher bodies centred around our physical body, and furthermore: "The hierarchy of realities is topped by the 'absolute'. The absolute is the basis of all realities. In its nonmanifest form, it is potential, intelligent energy. When rippled or modulated, it becomes the basis of

our tangible physical matter and individual matter"³. This model of philosophical caution provides the considerations that render Marks' "philosophical outlooks" somewhat "problematical".

Wagner admits that there are no scientific theories of paranormal investigation, but only because it "has as yet no paradigm (in the Kuhnian sense)". It is "not yet a mature science" being "in the pre-paradigm stage" and anyhow many would reject the "assumption that scientific thinking is necessarily analytical thinking". The desirability of scientific, analytical thought is thus "an open question".

RUSS FARRINGTON

33 Oakridge Road,
Bloomfield, New Jersey 07003, USA

1. Feyerabend, P. *Against Method* (New Left Books, London, 1975).
2. Feyerabend, P. in *The Journal of Philosophy* (Columbia University, 20 November 1969).
3. Bentov, I. *Stalking the Wild Pendulum* (Bantam, New York, 1971).

Preventing feticide

SIR—In a recent News article (*Nature* 324, 202; 1986), Radhakrishna Rao reports on the proposed government legislation to restrict the use of antenatal sex determination in India because of the growing problem of female feticide, but the problem may not be restricted to India. For some time now, we in the West Midlands have been aware of the problem of selective termination of chromosomally normal fetuses following amniocentesis for a legitimate purpose, for example to screen for Down's syndrome. It has been our practice (in common with most cytogenetic laboratories in the United Kingdom) to report the karyotypes of individual cases in full, including chromosomal sex. In order to prevent the possible abuse of a costly diagnostic procedure, we have, however, now decided to withhold sex from our reports, unless it has some direct clinical relevance (for example sex-linked disease prediction).

From 1 January 1987, it will be possible for an individual consultant to learn the fetal sex upon special request to the laboratory, but no longer will the chromosomal sex be routinely reported upon and freely displayed in the obstetric notes.

Perhaps the Indian government could make it similarly difficult to learn the chromosomal sex from written laboratory reports, rather than legislate against the useful diagnostic practice of antenatal chromosome analysis *per se*?

MAJ HULTEN
PAUL NEEDHAM

*Regional Cytogenetics Unit,
East Birmingham Hospital,*

JESSIE L. WATT
MIKE GRIFFITHS

*Regional Cytogenetics Unit,
Birmingham Maternity Hospital,
Birmingham B9 5ST, UK*

The domino effect explained

Diplomats should be glad to know that the bar-room game of letting dominoes knock each other over has a theoretical foundation at last. But there is much to be done before the theory will be satisfactory.

SINCE the flowering of the Cold War, diplomats everywhere have known what is meant by the domino effect; one country falls under pressure from an aggressor and, in doing so, knocks over its nearest neighbour, which knocks over the next, and so on. Not all diplomats may know that the phrase comes from the childish practice, a by-product of the game called dominoes, which is evidently itself a mediaeval poor man's card game, of standing small rectangular tiles on their smallest faces in such a way that each will knock over the next in line. Occasionally, the serious game is buried by the party trick that springs from it, as when television programmes show thousands of carefully prearranged dominoes falling to order.

Reprehensibly, all these exercises appear hitherto to have been based on a strictly empirical foundation. Even the latest attempt to provide a theory of the domino effect, by W.J. Stronge of the Engineering Department at the University of Cambridge (*Proc. R. Soc. A* **409**, 199–208; 1987), is applicable only to the simplest of all possible cases in which the dominoes stand and fall in a straight line.

Although Stronge's account of his work includes some comparison with experiment, nothing is said of when attention at the University of Cambridge will turn to the interesting cases, those when television producers try to persuade falling dominoes to generate a visually recognizable pattern. Attempts to simulate national flags seem to be especially popular, but advertising slogans are a possibility.

So far as it goes, Stronge's analysis is interesting enough. The starting assumption is that there is a one-sided infinity of dominoes standing vertically, with their broadest faces parallel to each other. The geometry of the problem is simplified by assuming that the spacing between consecutive dominoes is a constant, λ ; two other parameters matter, the width of the base on which the domino stands, which is really the thickness of the tile, h , and the height of the domino, L , which determines the position at which a falling domino will strike the next.

Stronge's interest is to find the conditions under which an impulse strong enough to fell the first domino will send all the others falling or, not quite the same thing, to tell when and how successive collapses will propagate as a wave along the chain. One way of tackling this problem

is to require that the kinetic energy of a falling domino, at the instant it hits the next, should not be less than the kinetic energy of the second when that, in turn, hits its successor. The kinetic energy of a falling domino by the time it hits the next can be calculated from the impulse derived from its predecessor and from the potential energy added by the decreased height of its centre of gravity.

Two high-school complications enter. The collision between one domino and the next will not be perfectly elastic, which means that the impulse acquired by a falling domino will be less than that transferred to the next in line. The usual statement is that the differences before and after impact of the velocities of the falling and struck dominoes, measured perpendicular to the face of the target, are in the ratio of the coefficient of restitution, e , which is a number less than unity unless the collision is elastic. Another complication is that of friction, a thief of energy at each collision as the impacting edge of a falling domino slides across its target.

The upshot of these calculations is that, for each geometrical arrangement of the dominoes, there is a characteristic speed of propagation of catastrophe determined by a characteristic velocity of impact of a falling domino with its successor. Moreover, if the object at the beginning of the line is set moving more quickly than required, extra energy will be dissipated until the characteristic speed is reached. The same is true if there is too little energy to begin with.

Stronge's chief result, a formula relating the characteristic impact velocity in a domino array to its geometrical properties and the dynamical characteristics of the objects, e and the coefficient of friction, μ , will unfortunately not be directly of assistance to those wishing to display the domino effect. The parameters are too deeply buried in the trigonometric functions. But what the inevitable numerical evaluations show is that, except when friction is important, the characteristic impact velocity increases with the distance between the dominoes (which is not surprising, for each then has further to fall before it hits the next).

The sensitivity of the result to the value of the coefficient of restitution, e , arises because the formula for the critical speed takes the form of an algebraic fraction whose denominator is $R^2 - (1 + e^2)$, where R is a quantity involving the coefficient of

friction and the geometry of the dominoes and whose value is numerically 2 when friction is literally zero. If e were also unity, meaning that no energy was lost in collisions, the characteristic impact speed would be infinity. That makes sense, for then the potential energy of all the fallen dominoes would be concentrated in that about to fall. No doubt special relativity would have to be invoked.

Unfortunately, Stronge has little to say about the speed with which catastrophe propagates along a line of dominoes. The reasons are understandable: the speed will be determined by μ and by the time that lapses between the point at which a domino is struck by its predecessor and that at which it hits the next in line, a quantity essentially that spent by a compound pendulum on part of a cycle, yielding an elliptic integral. But Stronge does provide three calculations of the speed of a wave of catastrophe in a line of realistic dominoes — giving values ranging between 0.65 and 0.86 m s⁻¹ as the spacing between the objects increases.

Quite apart from the restriction of this theory to linear arrays of dominoes, some awkward mechanical problems have also been overlooked. Thus the experiments at Cambridge, in which falling dominoes have been recorded by a cine camera, deny the assumption that each domino is knocked over by a single impact by its falling neighbour. It also seems that frictional losses may be especially important when dominoes are stood close together. Then the calculations depend critically on the assumption that the dominoes stand on a perfectly rough (slip-free) surface, which may not always be the case.

One general conclusion shining through these uncertainties is that, for a specified geometry, there must be a critical spacing below which no impulse, however great, will successfully propagate along the line. Physically, the energy will be absorbed in tilting dominoes so that they simply lean against each other.

It remains to be seen where this first comprehensive attempt at the theory of the falling domino will lead. Plainly there is an urgent need of more realistic calculations. Whether those with practical needs will be able to wait that long is another matter. But Stronge's theory is certain to stimulate a rash of empirical investigation in bar-rooms and other places where these investigations are habitually pursued.

John Maddox

Plant molecular biology

Transgenic cereals by direct injection of DNA into plants

Conrad Lichtenstein £30

CURRENTLY there are two successful approaches that can be used to transform plants genetically, but each method has limitations that are delaying the application of the techniques to certain commercially important crops. On page 274 of this issue¹, A de la Peña, H. Lörz and J. Schell report a significant step in applying a new method to cereals.

The first of the two current methods exploits a natural genetic engineer, *Agrobacterium tumefaciens*. This soil microbe contains a tumour-inducing (Ti) plasmid that transfers a DNA segment (the T-DNA) from the plasmid to the nuclear genome of infected plants (or *in vitro* to plant tissue). The method is restricted to dicotyledonous plants (dicots); monocotyledonous plants (monocots) are usually not susceptible (see my recent News and Views article² and refs 3 and 4 for reviews). See last week's issue⁵, however, for an exception.

The second method involves direct transfer of DNA to plant protoplasts, prepared by enzymatic digestion of cell walls, for example by chemically stimulated uptake using polyethylene glycol⁶ or a high-voltage pulse (electroporation), generating transient 'holes' in the protoplast membrane⁷. This method depends on a tissue culture system that allows regeneration of mature plants from protoplasts. Both methods use dominant selectable markers (for example, kanamycin resistance) to select for the transformed tissue which can then be screened for expression of co-transferred but unselected gene(s).

A major class of crops, the cereals, have so far evaded genetic manipulation. Because they are monocots they are not susceptible to agrobacterial infection and only in rice has it been possible to achieve plant regeneration from protoplasts. In this issue¹, de la Peña and co-workers describe a novel approach to genetic transformation of cereals obviating the need for *in vitro* tissue culture and regeneration. The authors injected DNA encoding kanamycin resistance into the developing floral side-shoots (tillers) of rye plants (*Secale cereale* L.) and demonstrated that some progeny had acquired kanamycin resistance by incorporation of this DNA.

Within the floral tillers are archesporial cells; these give rise to pollen in the developing pollen sac by two meiotic cell divisions. Previous work by de la Peña *et al.*⁸ had shown that these cells are permeable to colchicine at a stage 14 days before the first meiotic metaphase; colchicine

disrupts meiosis and thus provides an assay for uptake.

This result suggested that such cells might also be able to take up larger molecules such as DNA. Thus, de la Peña *et al.*¹ injected a plasmid encoding kanamycin resistance, and under the control of a promoter known to drive transcription in plants, into the floral tillers of about 100 rye plants. After cross-pollination with other similarly injected plants the seeds were harvested and germinated on medium containing kanamycin. Of about 3,000 seeds so germinated, seven seedlings remained green, indicating resistance to kanamycin — sensitive seedlings bleach and die. Of these seven, further biochemical analyses showed that two had aminoglycoside phosphotransferase activity (encoded by the kanamycin resistance gene). DNA hybridization studies confirmed that this expression results

from incorporation of the plasmid DNA. The plasmid DNA integrates in multiple copies, some of which show DNA rearrangements.

The selection method shows a background of kanamycin-resistant, yet untransformed, seedlings and the efficiency of transformation is low (about 0.07 per cent). A remaining question is whether unselected but co-injected DNA will integrate and express alongside the resistance gene at frequencies sufficient to allow detection by screening transformants.

Notwithstanding these provisos, one exciting prospect is that this simple transformation procedure can be applied to other cereals that currently lack transformation systems by DNA injection into similar pre-meiotic germline tissue. □

1. de la Peña, A., Lörz, H. & Schell, J. *Nature* **325**, 274–276 (1987).
2. Lichtenstein, C. *Nature* **322**, 682–683 (1986).
3. Lichtenstein, C. & Draper, J. in *DNA Cloning* Vol. II (ed. Glover, D. M.) Ch. 4 (IRL, Oxford, 1986).
4. Fraley, R. *et al. CRC Crit. Rev. Plant Sci.* **4**, 1–46 (1986).
5. Grimsley, N., Hohn, T., Davies, J. W. & Hohn, B. *Nature* **325**, 177–179 (1987).
6. Paszkowski, J. *et al. Meth. Enzym.* **118**, 668–684 (1986).
7. Fromm, M. *et al. Meth. Enzym.* (in the press).
8. de la Peña, A. *et al. Chromosoma* **83**, 241–248 (1981).

Conrad Lichtenstein is in the Centre for Biotechnology, Imperial College of Science and Technology, Prince Consort Road, London SW7 2AZ, UK.

Antigen recognition

Class discrimination in the world of immunology

Michael J. Bevan £35

THE upheaval which occurred in the early 1970s with the demonstration that helper and cytotoxic T lymphocytes recognize foreign antigens only in association with self surface glycoproteins encoded by the major histocompatibility complex (MHC) continues to be felt. Although we have grown used to the idea that helper T cells recognize degraded (peptide) forms of soluble protein antigens in the context of class II MHC products, we now have to cope with the realization that class I-restricted cytotoxic T lymphocytes, which are usually specific for cellular antigens such as viral proteins, also recognize degraded forms of the antigen.

It makes biological sense to direct lytic T cells towards the infected cell and to focus helper T cells on soluble antigen in order to increase the production of antibodies that will neutralize toxins or viral particles. But if both subsets recognize peptide fragments of foreign antigen, how is this distinction made?

It now seems that it is the responsibility of the two classes of MHC molecules to distinguish between soluble, endocytosed antigens and cellular antigens. The way this is accomplished, as discussed recently

by Germain in a News and Views article¹, may be by the non-random intracellular trafficking of vesicles that contain either class I or class II molecules. Vesicles that contain class II but not class I molecules intersect with the endocytic pathway and may present peptides derived from pinocytosed proteins. Class I-containing vesicles receive the peptide products derived from endogenously synthesized proteins. The latest information on the class I MHC-restricted recognition of integral cell membrane proteins is provided by two papers recently published in *Nature*^{2,3}.

Previous work on the class I-restricted cytotoxic T-lymphocyte recognition of cells expressing influenza nucleoprotein established that fragments of the protein were recognized^{4,5}. But this protein is made on free ribosomes. Thus, these experiments left open the question of whether integral membrane proteins, which are synthesized on membrane-bound ribosomes to be transported through the endoplasmic reticulum and Golgi complex and expressed in native form alongside class I molecules, would also be recognized as peptides by class I-restricted cytotoxic T lymphocytes. The

new work of Maryanski *et al.*¹ shows that when mice are immunized against syngeneic tumour cells that have been transfected with the fully functional gene for a human class I protein, cytotoxic T lymphocytes are generated that recognize a fragment of the human antigen in association with self class I molecules. The site on the human protein recognized by some class I-restricted T cells is defined by a stretch of 16 amino acids. The complete human class I antigen is simultaneously expressed at the surface, but this native form seems irrelevant for restricted recognition. The work of Townsend *et al.*², although not defining a peptide sequence of influenza haemagglutinin, shows that whether leader-sequence intact haemagglutinin is produced and expressed as an integral cell membrane component or whether leader-sequence deleted haemagglutinin is made in the cytosol and rapidly degraded, the same class I/haemagglutinin determinant is displayed at the surface for recognition. This means that the form of haemagglutinin recognized is denatured and implies strongly that it is a degraded form. The important message from this and other work is that irrespective of whether proteins are normally made in the cytosol or on the endoplasmic reticulum, they all provide degradation products that can associate on the cell surface with class I MHC molecules for the perusal of cytotoxic T lymphocytes.

This pathway of degradation of endogenously synthesized proteins leading to class I presentation seems to be different from that for the degradation of soluble exogenous proteins taken into the cell by endosomes and leading to class II presentation. For example, the latter pathway is sensitive to the drug chloroquine, whereas the former is resistant, and peptides derived from endogenously synthesized haemagglutinin might not interact with the class II pathway³.

Now that we are confronted with the fact that degradation products of endogenously produced proteins (whether originally destined for the nucleus, cytosol or membranes) can be presented on the cell surface with native class I molecules, we can ask whether class I presentation ever involves protein antigens that come from the exterior of the cell. This is central to understanding the induction phase of a

Wrinkle ridges on the Moon and the Earth

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

STUDY of the surfaces of Mars, Mercury and the Moon has revealed the existence of wrinkle ridges, complex geomorphic structures variously proposed to have volcanic, tectonic or combined volcanic/tectonic origins. Generally they consist of long linear ridges, with heights measured in metres, that may stretch for hundreds of kilometres. Such a lunar wrinkle ridge is shown (left) in a photograph of the southwestern part of the Mare Imbrium taken from Apollo 15; the field of view (about 370 km across) gives some idea of the scale of these features. Analysis of the ridges to determine their origins has been hampered by the absence of terrestrial data from which to extrapolate. J.B. Plescia and M.P. Golombek have recently investigated structures of Earth that they believe are the analogues of the planetary ridges; their report can be found in *Bulletin of the Geological Society of America* 97, 1289–1299 (1986). A typical terrestrial ridge of total length 37 km is shown (right), which formed after an earthquake at Meckering, Western Australia; the road and vehicles allow scaling. From this and other examples from many widespread localities, the authors argue that the ridges on Earth are the terrestrial equivalents of planetary wrinkle ridges. The major discrepancy is one of scale, in that the examples on Earth are considerably shorter; this is explained in terms of the relatively heterogeneous terrestrial stress systems, which effectively limit fault development. Planetary wrinkle ridges, it is suggested, developed following deformation associated with thrust faulting close to the planet surface. □

class I-restricted cytotoxic T-lymphocyte response to certain types of viral infections. If the only cell capable of presenting antigen to class I-restricted T cells is the infected cell itself, then sensitization would have to occur peripherally in the case of a virus that did not productively infect cells in the lymphoid organs. This is at odds with an old dogma that immune sensitization occurs centrally, in draining lymph nodes or in the spleen. There is, however, evidence in the transplantation literature suggesting that antigens can be picked up from the periphery and presented in the central lymph nodes. For example, female mice of a given H-2 type can be primed to reject a syngeneic male graft by cells from a male of a different H-2 type — presumably because the male-specific antigen is picked up and presented on host cells. In most cases, however, one does not know if the rejection mechanism is dependent on class I or class II MHC presentation. But from *in vivo* systems, there is clear evidence for such class I antigen presentation. Class I cytotoxic T lymphocytes can be primed to foreign minor antigens or to simian virus 40 T antigen that have been introduced on MHC-different cells^{7,8}. Furthermore, minor antigens may be picked up and presented by cells to tolerize class I-restricted T cells⁹. In these cases, the foreign anti-

gens must have been taken up by host cells and presented in association with their self class I molecules.

How can one explain the class I presentation of these exogenously derived cellular antigens? One way is to propose that the cell synthesizing the foreign antigen is also producing and leaking processed peptides that may bind class I molecules on cells of the lymphoid system. That this is feasible is demonstrated by the finding that synthetic peptides can be added externally to cells *in vitro*, thereby creating the cytotoxic T-lymphocyte target antigen^{3,4}. But it does not seem to be a likely occurrence *in vivo* where peptides are being flushed from the system, nor does it offer a basis for selecting for class I over class II presentation. Another more plausible way to take cellular antigens that are exogenous and to present them as endogenous, class I-associated antigens is via specialized antigen-presenting cells that phagocytose large cellular debris and shuttle the resulting peptide degradation products to their endogenous class I presenting system. Such a phagocytic cell may or may not express class II molecules.

The important distinction from the class II presentation systems here is that the exogenous antigens come as cell debris and not as soluble material or small particles which can be pinocytosed. Such a

1. Germain, R.N. *Nature* **322**, 687–689 (1986).
2. Townsend, A.R.M., Bastin, J., Gould, K. & Brownlee, G.G. *Nature* **324**, 575–577 (1986).
3. Maryanski, J.L., Pala, P., Corradin, G., Jordan, B.R. & Cerottini, J.-C. *Nature* **324**, 578–579 (1986).
4. Townsend, A.R.M. *et al.* *Cell* **44**, 959–968 (1986).
5. Townsend, A.R.M., Gotch, F.M. & Davey, J. *Cell* **42**, 457–467 (1985).
6. Morrison, L.A., Lukacher, A.E., Braciale, V.L., Fan, D.P. & Braciale, T.J. *J. exp. Med.* **163**, 903–921 (1986).
7. Fink, P.J., Weissman, I.L. & Bevan, M.J. *J. exp. Med.* **157**, 141–154 (1983).
8. Gooding, L.R. & Edwards, C.B. *J. Immun.* **124**, 1258–1262 (1980).
9. von Boehmer, H. & Hafén, K. *Nature* **320**, 626–628 (1986).

form of presentation would, of course, put the presenting cell at risk of being killed by class I-restricted cytotoxic T lymphocytes but, guarding against this, induced actively cytotoxic cells may move into the blood and lymph rather than stay in the central lymphoid organs. Are there two forms of presentation of exogenous antigen, one the endocytic pathway for soluble antigen

operative in all cells but immunologically relevant only in cells which express class II molecules, and another performed only by phagocytic cells? Time will tell. □

Michael J. Bevan is in the Department of Immunology, Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037, USA.

Structure determination

Solid-state NMR comes of age

Colin A. Fyfe

HIGH-RESOLUTION nuclear magnetic resonance spectroscopy (NMR) of liquids and solutions has developed into the most widely used analytical technique for the determination of the structures of unknown chemical compounds. It has recently become possible to obtain similar information from solids which, when used together with the information from traditional X-ray diffraction techniques, yields a complete picture of the structure of many materials. Elsewhere in this issue Kirkpatrick and co-workers (*Nature* **325**, 236; 1987) describe work on plagioclase feldspars that represents the state-of-the-art in this area, particularly because the complementary nature of the NMR and diffraction techniques is fully exploited.

In high-resolution solution NMR, signals are observed whose frequencies (chemical shifts) are diagnostic of the local magnetic (chemical) environment and, most importantly, whose relative intensities reflect the numbers of nuclei in the different environments. These two elements are critical in the determination of unknown chemical structures. However, NMR experiments on protons in solid materials yield only broad, featureless absorptions as the direct dipole-dipole interactions between the abundant ^1H nuclei are orders of magnitude greater than the chemical shifts from which structural information is obtained.

By working with dilute nuclei such as ^{13}C and ^{29}Si , and using high-power decoupling to remove dipolar interactions from the protons, 'magic-angle' spinning spectroscopy of the sample (Andrew, E.R. *et al.* *Nature* **182**, 1659; 1959 and Lowe, I.J. *Phys. Rev. Lett.* **22**, 285; 1959) to produce the isotropic chemical shifts and cross-polarization (Pines, A., Gibby, M.G. & Waugh, J.S. *J. chem. Phys.* **56**, 1776; 1972) to enhance the signal-to-noise ratio, J. Schaefer and E. Stejskal of Monsanto were able to obtain spectra of good resolution from completely solid samples (*J. Am. Chem. Soc.* **98**, 1031; 1976). In the case of many inorganic materials such as glasses, zeolites, minerals and clays, the interactions of protons are not important and the experiment is the very simple one

of magic-angle spectroscopy alone.

Suddenly, these experiments opened up the possibility of obtaining structural and chemical information on a range of materials and catalysed an upsurge of interest in many aspects of solid-state chemistry. The emphasis in most of these early studies was in extending the techniques and in applying them to different systems. Now, however, it is possible to see in more general terms how the early work fits in with the data provided by the more traditional diffraction investigations of solid systems.

In the case of amorphous materials such as surfaces, polymers and glasses, the NMR data are particularly important because diffraction studies yield extremely limited information as there is no long-range order in amorphous solids. Even in cases where single crystals are available, problems can arise where the structure is dynamic and/or disordered in nature or where individual atoms are difficult to distinguish (such as silicon and aluminium in minerals and synthetic zeolites) and the combination of the two techniques yields a more complete description of the structure. Because the NMR experiments provide the same information (isotropic shifts) in both phases, they can provide a bridge between phases, for example, the structure that exists in a single crystal as determined by X-ray diffraction techniques and the one that exists in solution.

The investigation of zeolites and minerals is a good example of the development of the combined use of NMR and diffraction data. The two techniques are complementary as NMR monitors the local environment (geometry of the atom, nature of the attached groups and local ordering) whereas diffraction data monitor long-range periodicity and lattice networks. Together, the two methods provide a complete description of the structure. Zeolites, open framework aluminosilicate structures, are widely used as catalysts, sorbents and molecular sieves in industry (Breck, D.W. *Zeolite Molecular Sieves* Wiley, New York, 1974). They are microcrystalline in nature and only limited powder diffraction data are available.

Because silicon and aluminium are adjacent in the periodic table, they have almost identical scattering factors, which makes it difficult to distinguish between them. In most systems, the atoms are disordered throughout the lattice. This distribution of local environments is exactly the kind of information which NMR can provide. E. Lippmaa *et al.* (*J. Am. Chem. Soc.* **103**, 4992; 1981) showed that the ^{29}Si magic-angle NMR spectra of these systems exhibit to a first approximation five peaks corresponding to the five local environments $\text{Si}[\text{}^4\text{Al}]$, $\text{Si}[\text{}^3\text{Al}, \text{Si}]$, $\text{Si}[\text{}^2\text{Al}, \text{}^2\text{Si}]$, $\text{Si}[\text{Al}, \text{}^3\text{Si}]$ and $\text{Si}[\text{}^4\text{Si}]$ describing the average distribution of silicon and aluminium throughout the lattice, and this has been confirmed and extended by others (Fyfe, C.A. *et al.* *Angew. Chem. Int. Edn* **22**, 259; 1983). There is a constant and additive effect on the shift for each aluminium atom added to the local environment. In addition, for highly siliceous systems where all environments are $\text{Si}[\text{}^4\text{Si}]$, signals are observed because of each of the crystallographically inequivalent sites in the unit cell (Fyfe, C.A. *et al.* *Chem. Lett.* **10**, 1547; 1983), and these shifts can be related to the local geometries of the atoms (see Smith, J.V. *et al.* *Nature* **303**, 223; 1983 and Thomas, J.M. *et al.* *Chem. Phys. Lett.* **102**, 158; 1983) and the effect of disorder can be described (Fyfe, C.A. *et al.* *J. Am. Chem. Soc.* **106**, 4435; 1984).

From these and related studies a clear set of general rules have emerged that can be used in extensions of the techniques to investigations of other aluminosilicate structures, most importantly minerals. The plagioclase feldspars described by Kirkpatrick *et al.* in this issue are the most abundant aluminosilicate minerals of the Earth's crust, but their detailed structures exhibit very complex and subtle variations and are quite imperfectly described, as is the distribution of silicon and aluminium atoms in all phases. The authors used ^{29}Si magic-angle NMR to investigate a series across the solid solution from ordered albite ($\text{NaAlSi}_3\text{O}_8$) to anorthite ($\text{CaAl}_2\text{Si}_2\text{O}_8$) at the two extrema. In the case of highly ordered albite, it is possible to assign the local environments related to the three peaks in the spectrum as is the case for these in the spectrum of a very special highly ordered anorthite. From these data it is possible to assign the peaks and interpret the distribution of local environments of the intermediate phases and to combine this information with that already available from diffraction data to propose mechanisms for the structural variations of the intermediate phases. Experiments of this type on disordered minerals and ceramics are currently under way in several laboratories. □

Colin A. Fyfe is in the Department of Chemistry and Biochemistry, University of Guelph, Guelph, Ontario N1G 2W1, Canada.

Neurobiology

Glia and the blood–brain barrier

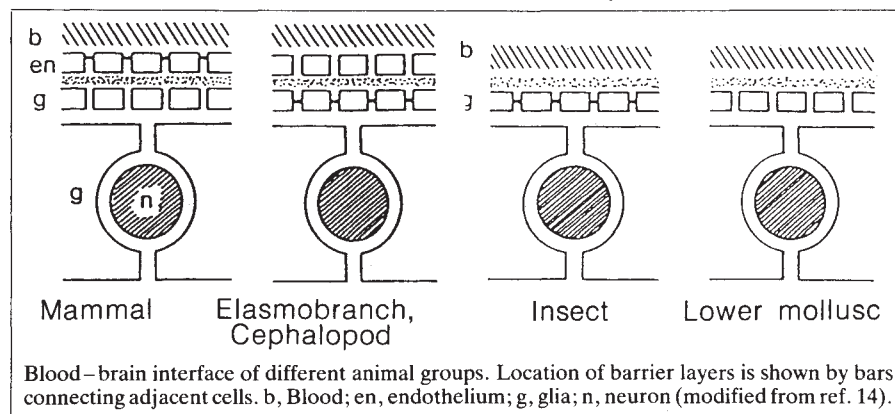
N. Joan Abbott

SINCE the classic dye experiments of Ehrlich¹ in 1885 it has been known that many substances injected into the blood have limited entry into the mammalian brain. Rival theories to explain this effect were a restricted extracellular space in the brain, or a barrier at the blood–brain interface. The proposed barrier was also variously attributed to the endothelium, the basement membrane or the end-feet of glial cells (astrocytes), which together form the brain microvessel wall. It was not until 1969, when Brightman and Reese² showed that the electron microscope tracer horseradish peroxidase was blocked at the level of endothelial tight junctions when injected into the blood or into the brain, that the debate was resolved. On page 253 of this issue³, R. C. Janzer and M. C. Raff provide clear evidence that the blood–brain barrier is induced by astrocytes. This result not only provides another function for astrocytes, but also could help to explain the breakdown of the blood–brain barrier that occurs in some pathologies of the brain⁴. However, comparison of the barrier layer in different animal groups suggests that the induction story will not be straightforward.

The role of perivascular glial cells has been tantalizingly difficult to establish. The well-ordered ‘pavement’ of glial end-feet on the endothelial wall suggested that these cells in some way influenced blood–brain exchange. There were doubts that the endothelial layer could account for the observed rate of bicarbonate transport at the barrier, and transport by the glial end-feet was suggested⁵. Orthogonal assemblies of particles are clustered on the endothelium-facing but not on the lateral membranes of astrocytes, indicating that the cells are polarized⁶. The functions of the assemblies remain obscure, although it is possible they are involved in ion transport. Neurons release potassium ions when they are active. Astrocytes have a high density of potassium channels on their end-feet⁷, providing a possible route for the ‘siphoning’ of potassium away from active neurons, so helping to maintain a stable neural environment.

Even before the location of the mammalian blood–brain barrier was known with certainty, Oldendorf⁸ in 1967 proposed that the barrier could be maintained by the influence of astrocytes on endothelial cells. Svendgaard *et al.*⁹ provided the first experimental evidence that brain tissue induces blood vessels to form a tight barrier by showing that two specific properties of the blood–brain barrier involving neurotransmitters, an enzymic trap-

ping mechanism for L-DOPA and a barrier for catecholamines, develop in brain fragments transplanted into the anterior chamber of the eye but not in iris fragments grafted into the brain. That is, the donor tissue and not the ingrowing vessels from the host control the vessel wall function. Stewart and Wiley¹⁰ were able to confirm this result in elegant grafting experiments between chick and quail embryos,



where they unequivocally demonstrated the origin of vascular elements from host tissue.

Janzer and Raff's new work clearly shows an inductive effect by astrocytes, both in the rat and the chick, where invading host blood vessels become relatively impermeable to albumin within an exogenous astrocytic mass. The use of primary cultures of identified type 1 astrocytes is a significant advance on earlier work. The next step will be to establish to what extent these induced barriers mimic the normal blood–brain barrier, for instance in impermeability to ions and presence of carrier mechanisms. Monolayers of brain endothelial cells seem to be capable of forming junctions tight to horseradish peroxidase even when cultured without astrocytes, but these junctions are still of relatively low resistance¹¹.

What is the nature of the inducing signal that Janzer and Raff demonstrate? Is it a diffusible substance, or an effect communicated by cell contact? The presence of tight junctions in blood vessels at the brain surface that lack an astrocytic sheath¹² suggests the inducer is diffusible and is secreted into the interstitial fluid and cerebrospinal fluid of the brain. Is it produced by isolated astrocytes, or by astrocytes only in the presence of endothelial cells? The model systems introduced by Janzer and Raff will be extremely useful for elucidation of the biochemistry of the interaction.

Some intriguing questions remain.

Many invertebrate groups have no blood–brain barrier. Insects, cephalopods and some arachnids do have a barrier, but it is glial, not endothelial^{13,14} (see figure). What is the inducing signal here, and how is the need for a barrier and its location decided? Even in mammalian brain, the barrier isolating parts of the nervous system from blood is not always at the endothelial layer — for instance it is found instead at the level of ependymal derivatives in the choroid plexus and over neurosecretory zones, and tight junctions between parenchymal glia may contribute to the isolation of neurosecretory regions from the rest of the brain¹⁴. Are these barriers induced by a different signal, or have the endothelial

cells here lost their sensitivity to the inducer? Among vertebrates, one anomalous group, elasmobranch fish, has its blood–brain barrier at the level of perivascular glial cells, although groups presumed close to the ancestral elasmobranch have an endothelial barrier¹⁵. What happened to the inducing signal or endothelial sensitivity in elasmobranch evolution? It seems certain that the induction of blood–brain barrier properties will turn out to be a complex and fascinating story. □

1. Ehrlich, P. translated in *Collected Studies in Immunity* (Wiley, New York, 1906).
2. Brightman, M.W. & Reese, T.S. *J. Cell Biol.* **40**, 648–677 (1969).
3. Janzer, R.C. & Raff, M.C. *Nature* **325**, 253–257 (1987).
4. Oldendorf, W.H. in *Cardiovascular Physiology, Microcirculation and Capillary Exchange* (eds Kovach, A.G.B., Hamar, J. & Szabo, L.) 349–353 (Pergamon, Oxford, 1981).
5. Pappenheimer, J.R. in *Capillary Permeability* (eds Crone, C. & Lassen, N.A.) 454–458 (Munksgaard, Copenhagen, 1970).
6. Landis, D.M. & Reese, T.S. *J. exp. Biol.* **95**, 35–48 (1981).
7. Newman, E.A. *Ann. N.Y. Acad. Sci.* (in the press).
8. Davson, H. & Oldendorf, W.H. *Proc. R. Soc. Med.* **60**, 326–328 (1967).
9. Svendgaard, N.-A., Björklund, A., Hardebo, J.-E. & Stenevi, U. *Nature* **255**, 334–337 (1975).
10. Stewart, P.A. & Wiley, M.J. *Dev. Biol.* **84**, 183–192 (1981).
11. Goldstein, G.W. *Ann. N.Y. Acad. Sci.* (in the press).
12. Bundgaard, M. *Brain Res.* **241**, 57–65 (1982).
13. Abbott, N.J., Lane, N.J. & Bundgaard, M. *Ann. N.Y. Acad. Sci.* (in the press).
14. Abbott, N.J., Bundgaard, M. & Cserr, H.F. in *The Blood–Brain Barrier in Health & Disease* (eds Suckling, A.J., Rumsby, M.G. & Bradbury, M.W.) 52–72 (Horwood, Chichester, 1985).
15. Cserr, H.F. & Bundgaard, M. *Am. J. Physiol.* **246**, R277–R288 (1984).

N. Joan Abbott is in the Department of Physiology, King's College London (KQC), Strand, London WC2R 2LS, UK.

Biology of communication

First squeaks of speech?

John C. Marshall

THE two hemispheres of the human brain are (roughly) symmetrical at the level of gross anatomy, yet their functional capacities are quite distinct; many domains of visuo-spatial cognition have their neuronal substrate in the right hemisphere, whereas the left is specialized for core aspects of language and speech processing¹. Although these latter capacities are species-specific properties of human-kind, we might nonetheless expect to find, in other animals, some of the underlying biological mechanisms upon which the evolution of language was built². In a study reported on page 249 of this issue³, G. Ehret, of the University of Konstanz, shows that the perception of communicative calls is lateralized to the left hemisphere in a comparatively lowly species — the house mouse.

A mouse pup that strays too far from the litter will emit ultrasonic calls within a narrow frequency range, and these sounds provoke the experienced mother to retrieve the pup to the nest. The mother is innately pre-wired to attend to ultrasonics, which are categorically centred on the frequency of 50 kHz, but must learn the significance of these 'distress' calls from interaction with her pups.

In Ehret's first experiment, the maternal motivation of primiparous lactating house mice was raised by removing pups from the litter and distributing them along a running board which extended to the right and left of a central

nest. After the mother had retrieved the pups and replaced them in the nest, two loudspeakers, each located at a different end of the running board, were switched on. One emitted 50-kHz and the other 20-kHz tone bursts. When tested binaurally, the lactating mice would leave the nest and reliably approach the 50-kHz ultrasound source in preference to the 20-kHz source. This preference persisted when the mothers' left ears were plugged to attenuate the signals by 40 dB, but each sound source was approached equally often when only the right ear was plugged.

Because, as in man, the auditory system of the mouse has primarily contralateral neuronal pathways from ears to cerebral hemispheres, the result demonstrates that the left hemisphere is specialized for some aspect of the complex task that the mice are here engaged upon. But which aspect? Perhaps the right hemisphere can 'hear' both 50- and 20-kHz tones but not discriminate between them. Perhaps it can so discriminate, but not link the discrimination to the respective sources of the tones. Perhaps the right hemisphere can succeed in both these steps, but does not 'know' that 50 kHz signifies distress (and thus does not 'care' which loudspeaker it goes to investigate).

Ehret's solution to this puzzle is extremely neat. In a second experiment, virgin females without experience of pups were placed in the same apparatus. For a reward of drinking water, they were trained to discriminate between 50- and 20-kHz tone bursts and to approach the source of the former tone in order to receive their reward. After conditioning to criterion with binaural presentation of the two simultaneous tone bursts, these animals were then likewise tested with either the left or the right ear plugged. The conditioned response to the 50-kHz tone persisted irrespective of which ear was attenuated.

When listening with either the left or the right ear, the mice can thus discriminate between the tones, locate the source of the preferred tone and respond appropriately by approaching that source. The cerebral hemispheres of the female house mouse seem equally adept in sound perception *per se*. It is only in the first experiment, where one of the sounds has an intrinsic biological significance (and that significance is known to the mouse) that the left hemisphere so dramatically outperforms the right. Furthermore, the 50-kHz distress call plays a crucial communicative role in the interaction of the mouse mother with her young. Ehret accordingly

speculates that an underlying lateral specialization for the auditory perception of communicative signals (between conspecifics) might be "a possible basis of the left-hemisphere advantage for speech sound recognition in man".

The evolutionary story must, however, be a good deal more complicated than this account suggests. In the human brain, it seems that the right hemisphere is specialized for the perception of affective aspects of communication, in both the visual and auditory modalities. Thus the recognition and interpretation of facial gesture⁴, and of the emotive significance of prosodic information in speech⁵, are more severely impaired by right than by left hemisphere damage. The perception of speech in terms of phonetic and phonemic categories is admittedly a strictly left-hemisphere function⁶, and one which is furthermore present in very young infants (as indexed by a right-ear advantage for the perception of speech sounds⁷). Yet there is good evidence that the right hemisphere can process speech in a non-phonological mode and may have semantic access to the meaning of auditorily presented words⁸. The linguistic specialization of the left hemisphere in man clearly extends well beyond the auditory domain; the mechanisms responsible for the expression and comprehension of gestural languages (such as American Sign Language) are as firmly committed to left-sided cortical areas as is any aspect of auditory communication⁹.

Nonetheless, Ehret has provided an important demonstration of a left-hemisphere advantage in an animal whose rank in the evolutionary scale might have led one to expect a functionally symmetrical brain. But why nature should choose an asymmetrical location for critical biological functions remains as mysterious as ever. The customary account which is given for lateralized control of human speech (or complex song production in canaries and chaffinches) — that it would be very difficult to ensure the requisite synchrony of timing in the operation of centres that were duplicated in two hemispheres — seems not to apply to the perception of 50-kHz tone bursts. □

100 years ago

A MOST extraordinary snowstorm occurred here today (January 7). In fifty years' experience I have seen nothing like it. It would be impossible to realize the gigantic size of the snowflakes without seeing them. I can only compare them to a fall of oranges, though the diameter of an orange would be small in com-



parison with thousands of these snowflakes. Snow had been falling with a slight thaw from 10 a.m., the snowflakes being small. Suddenly, at 12h 12m p.m. they became 2½ inches in length; at 12h 14m they had increased to 2¾ inches. At 12h 16m the flakes had increased in size to 3½ inches (and several measured were 4 inches across). Several flakes were sketched before they began to melt, and one of the sketches is sent as an illustration.

From *Nature* 35, 271; 20 January 1887.

1. Benson, D.F. & Zaidel, E. (eds) *The Dual Brain* (Guildford, New York, 1985).
2. Peterson, M.R. & Jusczyk, P.W. in *The Biology of Learning* (eds Marler, P. & Terrace, H.S.) 585 (Springer, Berlin, 1984).
3. Ehret, G. *Nature* 325, 249–251 (1987).
4. Campbell, R., Landis, T. & Regard, M. *Brain* 109, 509 (1986).
5. Ross, E.D. in *Localization in Neuropsychology* (ed. Kertesz, A.) 286 (Academic, New York, 1984).
6. Liberman, A.M. & Mattingly, I.G. *Cognition* 21, 1 (1985).
7. Mehler, J. in *Neonate Cognition* (eds Mehler, J. & Fox, R.) 7 (Erlbaum, London, 1985).
8. Duhamel, J.-R. & Poncet, M. *Neuropsychologia* 24, 769 (1986).
9. Marshall, J.C. *Nature* 322, 307 (1986).

John C. Marshall is in the Neuropsychology Unit, Neuroscience Group, The Radcliffe Infirmary, Oxford OX2 6HE, UK.

Planetary atmospheres

Clearer circulation on Uranus

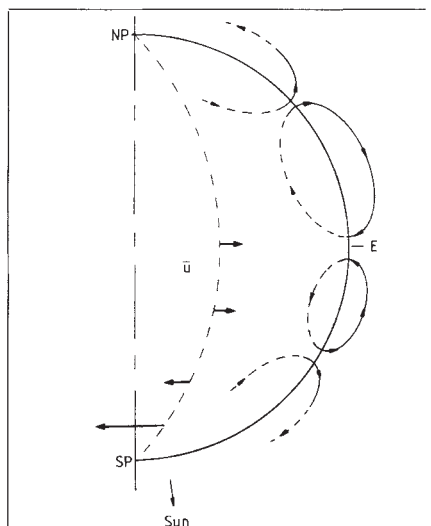
Peter L. Read

THE SPACECRAFT encounters by Voyagers 1 and 2 of Jupiter and Saturn in 1979 and 1981 yielded much new information on the upper-level circulations of their atmospheres. But the data were insufficient to resolve an earlier dilemma concerning the dominant energy source for the observed atmospheric motions. Both Jupiter and Saturn possess considerable interior sources of energy associated with continued gravitational contraction and differentiation which are capable in principle of driving an extremely deep-seated circulation. In addition, solar radiation affects a shallow surface layer. To unravel the dynamical implications of these two energy sources, some kind of controlled experiment was clearly desirable. One such opportunity was provided in January 1986 in the encounter by Voyager 2 of Uranus, a planet known to possess at most only a weak interior energy source, and whose solar heating would be strongly affected by the unique obliquity of its rotation vector to that of its orbit (98°). The initial results from this encounter (published as a special issue of *Science* **233**, 39–109; 1986) provide many new clues concerning the form of the circulation of Uranus, but also indicate several puzzling inconsistencies. From more detailed analyses of these measurements, presented at a recent meeting*, it is clear that several of these questions have been resolved, and a better picture of the atmospheric circulation of Uranus is beginning to emerge.

Preliminary infrared data had suggested scarcely any horizontal thermal contrast with latitude, despite the north pole having been in darkness for more than 40 Earth years. This had been anticipated¹ to result from the immensely long thermal inertia of the atmosphere, and was confirmed in some detailed radiative-convective calculations including seasonal effects (B. Bézard, Observatoire de Paris, Meudon). The solar-driven component of its circulation must therefore respond to the annual mean distribution of solar radiation, which is a maximum near both poles on Uranus, because of its obliquity. The simplest response to such a non-uniform distribution of heating is an axisymmetric circulation, with ascent in the region of strongest heating and descent where it is weakest, as occurs in the well-known Hadley circulation in the tropics of the Earth. For Uranus, however, this circulation should be in the opposite sense to that of the Earth (see figure), with the atmosphere descending at the equator.

Evidence in support of this form of circulation came from Voyager IRIS observations (B.J. Conrath, NASA Goddard Space Flight Center) showing small but vertically coherent temperature differences between one latitude circle and the next, which could be interpreted in terms of heating (or cooling) caused by adiabatic compression (or expansion) associated with downwards (upwards) vertical motion.

The transport of angular momentum by such a circulation would be expected^{2,3} to



Schematic cross-section of the solar-driven axisymmetric circulation on Uranus at the time of the Voyager 2 spacecraft encounter (NP, north pole; SP, south pole; E, equator; $\bar{u}$, cloud-top zonal flow with respect to the rotating planet). The main Hadley circulation comprises the two cells adjacent to the equator. Uranus rotates clockwise as seen from above the north pole.

have a strong bearing on the distribution of zonal (east–west) motion in the atmosphere of Uranus. Angular momentum would be transported polewards and vertically downwards by a reversed Hadley circulation, spinning up mid-latitudes relative to the tropics. The downwards vertical motion over the equator would deplete the upper (observable) atmosphere of angular momentum, resulting in a net tropical sub-rotation with respect to the rest of the atmosphere.

No further direct measurements of zonal motion were presented at the meeting beyond the lamentably few cloud-tracked wind measurements already available. Some remarkable indirect measurements of zonal motion near the equator have been made, however (G.F. Lindal, NASA Jet Propulsion Laboratory), using data from the Voyager radio occultation ex-

periment at latitudes between 2 and 7°S . By measuring the latitudinal curvature of constant pressure surfaces and comparing it with the local geoid, Lindal and colleagues could infer that the equatorial region of Uranus rotates more slowly than the magnetosphere (generally assumed to be tied to the electrically conducting deep interior), with a rotation period of about 18 hours (compared with the magnetic period of 17.24 hours). Thus, Uranus seems to be unique in the Solar System in possessing a retrograde equatorial jet near the top of its atmosphere. The sense of the geostrophic ‘thermal-wind’ shears^{5,6} in the tropics of Uranus, which are inferred from the horizontal temperature gradients, are therefore brought into line with similar measurements on Jupiter and Saturn, so that the pattern of zonal winds on all three major planets of the Solar System evidently decays with height above the visible clouds.

It seems inescapable that the solar-driven component is the dominant energy source driving the upper-level circulation of the atmosphere of Uranus. An assessment of the significance of this result for the other major planets, however, must await further refinements to measurements of the interior heat source of Uranus. It is reassuring to find that the observations of winds and temperature on Uranus (as on Jupiter and Saturn) can be interpreted within a coherent framework based on zonally symmetrical circulations, but several intriguing questions still remain.

The physical reason why all wind systems on the major planets decay with height above the visible cloud decks is still far from clear, although it probably entails the effects of momentum transfer by eddies. There is little direct evidence of such eddies in the atmosphere of Uranus, yet theoretical analyses^{4,5} of idealized prograde jets, such as found on the planet at mid-latitudes, suggest a strong propensity for such instabilities, much as in the troposphere of Earth. The presence of prograde flow with respect to the magnetosphere (and hence, by assumption, to the deep interior) at latitudes polewards of the equatorial-reversed Hadley circulation in each hemisphere probably indicates the presence of a further circulation cell at higher latitudes (as shown in the figure), although the mechanism for such an indirect circulation, with descending motion implied near the sub-solar point, is not yet apparent.

The circulation pattern inferred from the infrared data is evidently asymmetrical between the two hemispheres, in the sense that the entire pattern seems to be displaced along the rotation axis towards the (dark) north pole. This apparent displacement is in the same sense as that of the dipole component of the magnetic field of Uranus (F.M. Flasar, NASA God-

* Division of Planetary Sciences of the American Astronomical Society meeting, Paris, 4–7 November 1986.

dard Space Flight Center), the centre of which is located well away from the centre of the planet. Such a coincidence hints at the possibility, speculated on by D.J. Stevenson (Caltech), of magnetohydrodynamic coupling between surface motions and the electrically conducting interior (as also suggested⁶ for Jupiter and Saturn), although details of such an interaction remain to be explored. □

1. Stone, P.H. *Icarus* **24**, 292–298 (1975).
2. Read, P.L. *Q. J. R. Met. Soc.* **112**, 231–272 (1986).
3. James, I.N. *Weather* **41**, 21–28 (1986).
4. Pedlosky, J. *Geophysical Fluid Dynamics* (Springer, New York, 1979).
5. Gill, A.E. *Atmosphere-Ocean Dynamics* (Academic, London, 1982).
6. Kirk, R.L. & Stevenson, D.J. *Bull. Am. Astr. Soc.* **15**, 837 (1983).

Peter L. Read is in the Geophysical Fluid Dynamics Laboratory, Meteorological Office, London Road, Bracknell RG12 2SZ, UK.

Nomadic agriculture

Mobile resources for survival

Peter D. Moore

NOMADIC pastoralism in the semi-arid regions of the world is a hazardous occupation because of the unpredictability of water supply and therefore the unreliability of forage production¹. The vegetation of such areas is often patchy² but contains a range of species that vary in their drought tolerance and in their palatability to herbivores³. Under such circumstances it could be advantageous for pastoralists to maintain mixed flocks of grazing animals so that resource use can be kept at a maximum even when those resources are depleted through drought. Coppock, Ellis and Swift^{4,5} have recently examined this proposal by studying the food intake of the various components of mixed flocks of animals in Kenya and conclude that the use of several species, combined with the mobility inherent in the nomadic way of life, does indeed enhance the efficiency of grazing systems and increase the likelihood of survival in times of drought.

Vegetation in semi-arid lands varies both in space and time. Plant biomass and productivity generally increase with water supply and a range of plant forms may contribute to this, including shrubs (or even trees like *Acacia*), perennial grasses, bulbous perennials, and annual grasses and herbs. Some of these, particularly the annuals, may be available for grazing only for a short period each year during the wet season so that variability occurs in time as well as in space⁶. This fluctuating plant biomass forms the resource base on which pastoralism is founded.

The dietary preferences of the animals used by pastoralists can be studied by analysing the gut contents of dead animals⁷, by surgically examining the living (fistulation), by fecal analysis⁸ or by detailed observation of grazing behaviour. Analysis of the gut contents is wasteful and fistulation involves many practical problems in a nomadic herd. There are also several problems with fecal analysis, including differential digestion and the problems of identification when dealing with a diverse array of plant cuticles, hairs and so on^{9,10}. The final option, observation, requires the logging of the number of bites a parti-

cular animal takes from each plant species in a given time, together with the determination of bite size for that animal. Measuring bite frequency and at the same time identifying the prey plant can be quite difficult, especially with some fast eaters; many grazers, like sheep, cattle, deer and kangaroo have a rate of about 50 bites per minute, but some speedier ungulates,



Good mixers: sheep flock in north-east Iran.

such as reindeer, can exceed 200 (ref. 11). It takes a sharp eye to identify lichens at that speed.

Using this method of diet analysis it is possible to establish how domestic grazers exert their preferences and thus partition the resources of the environment between them¹². Nyerges¹³, for example, studied the grazing behaviour of sheep and goats in the semi-arid scrub of eastern Iran and found that although the demands of the two animals are indistinguishable as far as herbaceous materials are concerned, there is little overlap in their browsing activities on the shrubs. Goats exploit the *Pteropryum* and *Amygdalus* scrub in the dried river beds whereas sheep are prepared to consume saline-tolerant succulents like *Salsola*. The one overlap is the small shrub *Artemisia herba-alba*, which is eaten by both sheep and goats (17 per cent of goat diet and 23 per cent of sheep diet).

The studies of Coppock *et al.*^{4,5} were conducted in the arid northwestern region of the Kenyan Rift Valley among savanna vegetation, where the average net annual

above-ground productivity is 167 g m⁻², which compares poorly with the 1,071 g m⁻² from the grasslands of the Nairobi National Park¹⁴. The nomadic people of Ngisonyoka Turkana use several different domesticated grazers to exploit this poor productivity. Cattle are specialists in grass consumption (96 per cent of diet), whereas at the other end of the spectrum camels mainly browse on shrubs and trees (95 per cent). Between these extremes, sheep, goats and donkeys enjoy a more mixed diet.

The different types of forage vary in their nutritional value throughout the year, grasses becoming particularly poor in value during the dry season. This means that cattle fail to obtain green fodder within a month of the end of the short wet season. Camels, on the other hand, maintain some green input in their diet throughout the year by feeding on shrubs and trees with deeper roots and hence a more reliable water supply⁵. The more versatile habits of sheep and goats also permit some variety in diet as they can switch from herbs to shrubs when the former become more scarce or less nutritious. Overall, cattle consume more fibre whereas camels select for materials higher in protein and dietary cell solubles, but this leaves them with a relatively low energy intake.

The specialization of diet, both spatial and temporal, observed in these animals is reminiscent of that seen in the natural grazers of the area^{15,16} and is clearly well adapted to maximize use of resources. Coppock *et al.* point out that there are considerable advantages in maintaining this diversity of grazing animals used in the traditional pastoral systems rather than encouraging a concentration on a single species. Equally, the mobility which is an inherent feature of the nomadic system should be allowed to continue in order to provide opportunities for spatial as well as temporal variations in diet. □

1. Schwartz, H.J. *Impact Sci. Society* **30**, 279–288 (1980).
2. Larmuth, J. & Harvey, H.J. *J. Arid Environ.* **1**, 129–133 (1978).
3. Coe, M.J., Cumming, D. & Phillipson, J. *Oecologia* **22**, 341–354 (1976).
4. Coppock, D.L., Ellis, J.E. & Swift, D.M. *J. appl. Ecol.* **23**, 573–583 (1986).
5. Coppock, D.L., Swift, D.M. & Ellis, J.E. *J. appl. Ecol.* **23**, 585–595 (1986).
6. Walker, B.H., Ludwig, D., Holling, C.S. & Peterman, R.M. *J. Ecol.* **69**, 473–498 (1981).
7. Baharav, D. *J. Arid Environ.* **4**, 63–69 (1981).
8. Bhadresa, R. in *Methods in Plant Ecology* (eds Moore, P.D. & Chapman, S.B.) 61–71 (Blackwell Scientific, Oxford, 1986).
9. Stewart, D.R.M. *Bot. J.* **84**, 63–174 (1965).
10. Stewart, D.R.M. *J. appl. Ecol.* **4**, 83–111 (1967).
11. Short, J. *J. appl. Ecol.* **23**, 559–571 (1986).
12. Coughenour, M.B. *et al. Science* **230**, 619–625 (1985).
13. Nyerges, A.E. *ODI Pastoral Network Pap.* **9d**, 1–44 (1979).
14. Deshmukh, I. *J. appl. Ecol.* **23**, 115–123 (1986).
15. Gwynne, M.D. & Bell, R.H.V. *Nature* **220**, 390–393 (1968).
16. Leuthold, W. *Oecologia* **35**, 241–252 (1978).

Peter D. Moore is in the Biology Department, King's College London (KQC), Campden Hill Road, London W8 7AH, UK.

Number theory

Geometry finds factors faster

Ian Stewart

EVERY whole number can, in theory, be resolved into prime factors. In practice the process may take longer than the lifetime of the Universe. The obvious method, trial division by all primes up to the square root, is hopelessly inefficient. Its execution time grows so rapidly that improvements of a factor of, say, 100 in computer speeds will add only a few digits to the size of numbers that can be dealt with. To factorize an arbitrary 250-digit number — that is, one without special features that make it easy to factorize — is currently out of the question.

On the other hand, nobody has ever proved that prime factorization really must be as hard as it seems to be. Efficient and effective methods have not been ruled out. A new method, devised by Hendrik Lenstra of the University of Amsterdam and based on deep ideas from algebraic geometry and number theory, points to a new direction for attack (*Factoring Integers with Elliptic Curves*, Mathematical Sciences Research Institute preprint, Berkeley, 1986). The problem is of practical importance because of its relation to public-key cryptosystems — methods, hoped to be 'unbreakable', for putting messages into code. If a fast-factor algorithm exists, some of these methods will be breakable. The problem also has intrinsic mathematical interest as one of the most basic questions in number theory.

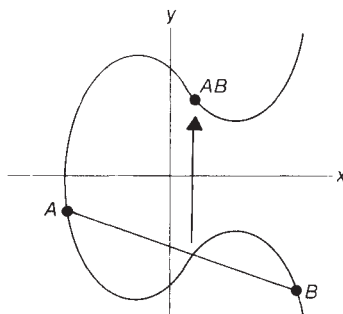
The objective is an algorithm: a prescribed procedure, involving no guesswork, and guaranteed either to produce a factor or to show that the number is prime. The efficiency of an algorithm is measured by the way its worst-case 'running time' grows with the number n of digits of the number being factorized. The running time is usually defined as the number of arithmetical operations ($+$, $\times$ and so on) involved, and only a relatively crude estimate is sought. A 'good' algorithm runs in polynomial time, growing no faster than some fixed power of n ; a bad one runs in exponential time, growing as the n th power of some constant.

Erratum

In the article 'Melting and the surface' by Robert Cahn (*Nature* **323**, 668; 1986) a sentence in the sixth paragraph was omitted. The second half of the paragraph should have read: "The melting temperature of argon at atmospheric pressure is 84 K, but allowing for the pressure in the bubbles (estimated from the measured anomalously small lattice parameter of the argon) it should be about 250 K. In fact, the diffraction pattern of solid argon is detectable up to 730 K, that is, a metastable superheating of 480 K!"

For an n -digit number of size roughly 10^n , trial division requires about $10^{n/2}$ trials. Even at one arithmetic operation per trial (which is far too low an estimate) this has exponential growth, hence is inefficient. In contrast, the euclidean algorithm for finding the least common multiple (l.c.m.) of two n -digit numbers takes about $5n$ steps. So this runs in polynomial time and is highly efficient.

To illustrate the alternatives to trial division, I will describe Pollard's $p-1$



The group law on an elliptic curve. Given two points A and B , form the line joining them. This cuts the curve in a third point. Reflect that point in the horizontal (y) axis to get the point AB , which is the 'product' of the original points A and B .

method (see Riesel, H. *Prime Numbers and Computer Methods for Factorization*, Birkhäuser, Boston, 1985). This method starts with a number N and looks for prime factors p such that $p-1$ has only 'small' prime divisors. It is based on a number-theoretic fact: if $p-1$ divides a number Q and p does not divide Q then p divides $a^Q - 1$ for any a not divisible by p . Then the l.c.m. of N and $a^Q - 1$ is divisible by p , and this can be found rapidly by the euclidean algorithm. In practical implementations a standard series of Q s is defined and stored in the computer: for example, $Q = \text{l.c.m.}(1, 2, \dots, r)$ for $r = 1, 2, 3, \dots$ up to some predetermined limit. As an example, the method has been used to find the factor $p = 2670091735108484737$ of $3^{136} + 1$. Although p is large, $p-1$ has the factorization $2^7 \cdot 3^7 \cdot 7^2 \cdot 17^2 \cdot 19 \cdot 569 \cdot 631 \cdot 23993$, and the primes involved there are all small.

The key to Lenstra's method is to view Pollard's algorithm in more abstract terms. Suppose that p is a prime. If any two of the set of numbers $1, 2, \dots, p-1$ are multiplied together modulo p (that is, the remainder is taken on dividing the product by p) then another number in that set is obtained. The set is said to form a group under multiplication modulo p . Pollard's method can be seen as exploiting the structure of this group. Lenstra's method

is obtained from Pollard's by replacing the group by a more subtle one: the group of points on an elliptic curve.

Elliptic curves have been studied by number-theorists for about a century; not for applications to prime factorization, but because of their intrinsic mathematical beauty and interest. They are curves defined by an equation of the form $y^2 = x^3 + ax + b$, for constants a and b . Associated with any such curve is a geometric multiplication law. Given any two points A and B on the curve, draw the line between them (see figure). Because the curve has an equation of degree 3 this line meets the curve at a third point. Multiply the y -coordinate of this third point by -1 : because of the term y^2 in the equation, the result is still a point on the curve. Call it the 'product' AB of the first two points. The surprise is that this product obeys sensible algebraic laws: the points on the curve form a group.

In Lenstra's algorithm the coordinates x and y are taken to be integers modulo N , where N is the number to be factorized. If a single elliptic curve is used the method is much like Pollard's. However, the number $p-1$ that occurs in Pollard's method is replaced by the number $p-t$ where t depends on the choice of elliptic curve. Different curves give different t s, so there is a whole range of possible numbers $p-t$. As long as just one of these has small divisors, the method will find a factor.

The algorithm therefore starts by choosing some elliptic curve and seeking a factor by a generalized version of Pollard's method. If this fails, it does something that is not available in the original Pollard method: it tries another elliptic curve. Provided there is a reasonably dense set of numbers near p that are a product of small primes, the method will find a factor rapidly. The precise running time is rather complicated, and its proof depends on a plausible but unproved conjecture on this density. Other known methods have a similar conjectured running time, but they lack one important feature of Lenstra's method: the running time depends on the size of the prime factors of N .

Until very recently nobody had expected there to be any connection between elliptic curves and prime factorization. Only with hindsight is the connection clear. Although Lenstra's method may seem complicated, it is a natural generalization of Pollard's method. Perhaps even more efficient primality tests lurk unsuspected among the discoveries of number theorists. If so, some deep and abstract mathematics will take on a new, practical guise. Computer scientists working on algorithms for factorization would be well advised to brush up on their number theory. □

Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Evolutionary adaptations

Aristotle's theory of mammalian teat number is confirmed

Jared M. Diamond

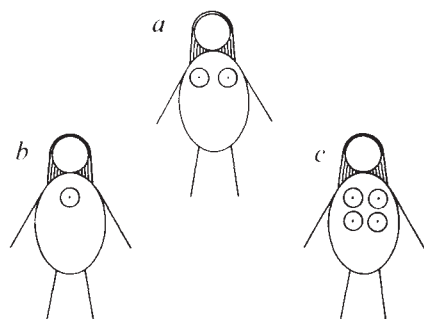
Animals which have small litters, both those that have solid hoofs and those that carry horns, have their mammae by the thighs; and there are two of them. Animals that have large litters or are polydactylous . . . have numerous mammae placed at the sides upon the abdomen — for example, swine and dogs. (Aristotle, *The Parts of Animals* Lesbos, about 350 BC).

Le nombre des mamelles et leur situation sont extrêmement variables . . . Il est d'ailleurs ordinairement en rapport avec le nombre des petits que les femelles peuvent mettre bas. (G. Cuvier, *Leçons d'Anatomie Comparée* vol. 5, p.155; Baudouin, Paris, 1805).

MAMMARY glands are so central to ecology and life history in the order Mammalia that they give the order its name. We think anthropocentrically of the normal teat number as two but, as Aristotle noted, it varies greatly among species: some primates have six, and rodents have up to fourteen. What accounts for this interspecific variation? Why, for instance, does the human female have design *a* rather than *b* or *c* (see figure)?

Aristotle suggested that teat number is related to litter size, a view followed by Cuvier but variously espoused or disputed by twentieth century scientists. A recent analysis by A.N. Gilbert (*Proc. natn. Acad. Sci. U.S.A.* **83**, 4828–4830; 1986), resting on a massive database, now supports Aristotle. Gilbert assembled values, for 266 rodent species, of teat number *T* (ranging from 2 to 14), mean litter size $\bar{L}$ (ranging from 1 to 8) and maximum litter size $L_{\max}$. There is a significant positive correlation between $\bar{L}$ and *T*: $\bar{L} = 0.39 \pm 0.46T$, $P < 0.001$, $r = 0.72$. Averaged over all species, the ratio $\bar{L}/T$ is 0.53 ± 0.01 , and $L_{\max}/T$ is 0.92 ± 0.02 . Thus, there are two teats per pup for an average litter, one for a maximal litter.

This observed correlation begs the question as to cause and effect: does litter size determine teat number or vice versa?



Theoretical designs for human mammary gland arrangement. *a*, The accepted number; *b*, no excess capacity for twins; *c*, too extravagant.

At a proximate level, teat number does affect surviving litter size, as milk production is critical to pup survival and there is a sharp increase in pup mortality when litter size exceeds teat number. The value $L_{\max}/T = 0.92$ means that natural selection has produced mammals with a safety margin of teats, enough to provide one per pup in the occasional large litter.

At an ultimate level, there have been many theoretical discussions of the evolution of litter size, which is generally seen as constrained by factors such as latitude, length of breeding season, body size and metabolic rate. By ignoring teat number, such analyses tacitly assume that it easily evolves in response to the driving variable

of litter size. However, Gilbert suggests that this may not be true. For instance, within a few generations it is feasible to select animals for increased litter size but not for increased teat number, which is a rather constant species characteristic.

If we now return to our question concerning human females, they are seen to fit Gilbert's rodent-based theories perfectly. For $T = 2$, the relations $\bar{L} = 0.53T$ and $L_{\max} = 0.92T$ predict that our maximum litter size should be 1.84 (after rounding to an integer, twins) and our mean litter size 1.06 (usually singletons). In fact, singletons are our norm, twins a not uncommon exception (5 per cent of all births in some populations) and triplets vanishingly rare. Thus, our selected design *a* in the figure is to accommodate the occasional twin birth, whereas the other designs are rejected because they lack a safety margin (*b*) or are extravagant (*c*). □

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Edward A. Doisy (1893–1986)

EDWARD A. DOISY died on 23 October 1986, one month before his 93rd birthday. His numerous scientific achievements centred around his masterful selection of methods for the isolation of trace natural products of biological significance. Needless to say, the methods available to Ed Doisy in the early part of this century were crude compared to those available today. His major contributions were made while residing in the city of St Louis, where he worked first at Washington University (1919–1923) and then at St Louis University (1923–1965).

Edward Doisy was born in Hume, Illinois, on 13 November, 1893. He studied chemistry at the University of Illinois and then went to Harvard to work under the guidance of Otto Folin, receiving his doctoral degree in 1920.

He was appointed an instructor in biochemistry at Washington University School of Medicine in 1919 and became associate professor in 1923, when he moved across town to become the first professor and chairman of a new department of biochemistry at St Louis University, a position he held until his retirement in 1965.

At Washington University he began his work on the isolation of natural products by the purification of insulin by isoelectric precipitation. He then developed an interest in ovarian function and the nature of the ovarian hormone through collaborative studies with the anatomist Edgar Allen. By judicious application of bioassays for oestrus in rats and his skill in the purification of trace compounds, he was the first to obtain oestrone, one of the female sex hormones, in crystalline form from urine in 1929. A few years later, he

isolated from several tons of sow ovaries the more potent ovarian hormone oestradiol. These discoveries opened up the field of fertility to biochemical study, the consequences of which are still being exploited.

In 1936, following the discovery of vitamin K as a new fat-soluble compound that promoted coagulation of blood in chicks and rats by Henrik Dam, Doisy undertook the isolation of this lipid vitamin from various sources, including alfalfa and fermented fish meal. In 1939, he succeeded in crystallizing vitamin K from alfalfa meal and vitamin K₂ (now called menaquinone-7) from fermented fish meal. He subsequently synthesized homologues in both the vitamin K₁ and K₂ series. Doisy received the Nobel Prize in physiology and medicine, jointly with Dam, the discoverer of vitamin K₁, in 1943 for his work on the isolation and synthesis of vitamins K₁ and K₂. No doubt his contributions to the oestrogen field played some role in the award.

During World War II, he worked for the Office of Scientific Research and Development and isolated several new antibiotics. After the end of the war, Doisy began to study bile acids and succeeded in isolating novel bile acids from the rat.

His many students and associates remember Edward Doisy as a single-minded and astute judge of research problems that could be solved with the techniques available, and an excellent mentor who both recognized scientific skills in young people and helped to develop them to the utmost.

Robert E. Olson

Robert E. Olson is Professor of Medicine at the State University of New York at Stony Brook, Stony Brook, New York 11794-8160, USA.

Bone nitrogen isotope composition and climate

SIR—Heaton *et al.*¹ report that the $^{15}\text{N}/^{14}\text{N}$ ratios of bone collagen from prehistoric humans with little or no access to marine protein show a negative correlation with present rainfall levels in the areas of South Africa they inhabited. Such a correlation cannot, however, be accepted uncritically, because the influences of dietary differences among such populations have not been considered. For example, $^{15}\text{N}/^{14}\text{N}$ ratios of bone collagen of early historic Griqua pastoralists from the arid interior of Orange Free State are higher than those of late prehistoric Iron Age farmers of the Northern Transvaal². This is expected given the well-known trophic level effect on nitrogen isotope ratios of bone collagen³. The data for these two populations, if plotted together against rainfall, would produce a regression similar to that obtained by Heaton *et al.*, but would not reflect only climatic influences. Thus the authors' interpretation of the nitrogen isotope ratios for human bone collagen based strictly on climate seems to be premature.

By contrast, their data for modern herbivores demonstrate a correlation between rainfall levels and the nitrogen isotope ratios of bone collagen that cannot be ascribed to dietary differences because no relationship between plant $^{15}\text{N}/^{14}\text{N}$ ratios and rainfall was observed¹. The authors suggest that the higher ratios of "mammals in arid areas are therefore more likely to be linked with the nitrogen metabolism in the body of the animal itself". In the absence of more specific explanations we feel compelled to summarize our own model and correct a potential flaw in its original presentation⁴.

The enrichment of ^{15}N in the bone collagen of East African water-conserving herbivorous mammals relative to water-dependent ones that we reported suggested a relationship between physiological mechanisms of water conservation and nitrogen isotopic mass balance⁴, as follows. Urea, the major form in which nitrogen is excreted by mammals, is known to perform an essential function in urinary water conservation⁵. Herbivorous mammals on diets high in protein have the capacity to excrete a highly concentrated urine under conditions of heat and water stress. This is accompanied in most species by an absolute and sometimes spectacular increase in urea output, as in the dikdik⁶ and two cattle breeds⁷, though not in the camel⁸ or in cattle on low-protein diets⁷. Urine in *Bos taurus* has been shown to be depleted in ^{15}N relative to the diet, with the depletion being greater during the day than at night⁹ (when water stress is presumably lower). If this difference occurs in all mammal species, animals under more continuous water stress would re-

spond by excreting a more concentrated urine with a quantitative increase in the excretion of ^{15}N -depleted urea, resulting in higher $^{15}\text{N}/^{14}\text{N}$ ratios in the unexcreted nitrogen. In order to satisfy mass balance constraints, the tissues synthesized from this remaining nitrogen pool in water-conserving, urea-excreting mammals should have higher $^{15}\text{N}/^{14}\text{N}$ ratios than those of mammals that do not conserve water and/or intensively recycle urea. Such a mechanism would explain the nearly one-to-one correlation in rank order we have found between maximal urinary osmolality⁷ and mean bone collagen $^{15}\text{N}/^{14}\text{N}$ ratios for zebu cattle, donkey, impala, goat, sheep and dikdik⁴, and most of the data for herbivores presented by Heaton *et al.*¹.

This model is only a hypothesis but it can be tested by controlled water-stress experiments and analysis of nitrogen isotopic mass balance in animals living in different climates and microhabitats. Until the contributions of ecological and physiological processes to variation in animal

tissue nitrogen isotope ratios are more clearly understood, palaeodietary analysis based on $^{15}\text{N}/^{14}\text{N}$ ratios of bone collagen must be regarded as uncertain in all but the simplest of ecological contexts.

STANLEY H. AMBROSE

Department of Anthropology
University of Illinois,
Urbana, Illinois 61801, USA

MICHAEL J. DENIRO

Department of Earth and Space Sciences,
University of California,
Los Angeles, California 90024, USA

1. Heaton, T.H.E., Vogel, J.C., von la Chevallerie, G. & Collett, G. *Nature* **322**, 822–823 (1986).
2. Schoeninger, M.J. & DeNiro, M.J. *Geochim. cosmochim. Acta* **48**, 625–639 (1984).
3. Ambrose, S.H. & DeNiro, M.J. *Nature* **319**, 321–324 (1986).
4. Ambrose, S.H. & DeNiro, M.J. *Oecologia* **69**, 395–406 (1986).
5. Beeekes, R. in *A Companion to Animal Physiology* (eds Taylor, C.R., Johansen, K. & Bolis, L.) 266–288 (Cambridge University Press, 1982).
6. Maloiy, G.M.O. *Phil. Trans. R. Soc. B* **184**, 167–178 (1973).
7. Livingstone, H.G., Payne, W.J.A. & Friend, M.T. *Nature* **194**, 1057–1058 (1962).
8. Maloiy, G.M.O. *Symp. zool. Soc. Lond.* **31**, 243–259 (1972).
9. Steele, K.W. & Daniel, R.M. *J. agric. Sci. Camb.* **90**, 7–9 (1978).

Signal-transduction

SIR—Wakelam *et al.*¹, as discussed in News and Views by Petersen and Bear², have demonstrated that although concentrations of glucagon above 10^{-8} stimulate the intracellular concentration of cAMP but not of inositol phosphates in hepatocytes, low concentrations (10^{-9}M) increase inositol phosphates but not cAMP. We would like to draw attention to the corresponding effects of ACTH on cells from the rat whole adrenal cortex; concentrations of ACTH above 10^{-10}M stimulate cAMP levels but neither ^{32}P incorporation into phosphatidic acid and phosphatidylinositol nor the production of inositol phosphates, whereas $\sim 10^{-12}\text{M}$ ACTH produces the opposite results^{3,4}. Moreover, we have obtained similar results using rat zona fasciculata-reticularis (ZF-ZR) adrenocortical cells^{5,6} and have shown that high concentrations of ACTH inhibit the increase in production of inositol phosphates by angiotensin II in ZF-ZR cells⁷.

Wakelam *et al.*¹ also report that TH-glucagon, an agonist for all the biological effects of glucagon, is effective in stimulating the production of inositol phosphates but not the intracellular concentrations of cAMP. There are also ACTH analogues that are effective in stimulating steroidogenesis in ZF-ZR cells but are only weakly active in activating adenylyl cyclase⁸. Their activity in stimulating phospholipase C would now be of interest.

The steroid output of ZF-ZR cells is stimulated mainly by ACTH-like peptides but zona glomerulosa (ZG) cells of the adrenal cortex, which produce aldosterone, are stimulated by many different

factors. These factors employ either the adenylyl cyclase or phospholipase C system⁹. There are as yet no reports of ZG stimulators acting through both systems in the concentration-dependent manner of ACTH in ZF-ZR cells. This would be an appropriate area for further study.

GUY ST J. WHITLEY

Department of Immunology,
St George's Hospital Medical School,
London SW17 0RE, UK

JAMES F. TAIT

Moorlands, Main Road,
East Boldre, Hants SO42 7WT, UK

1. Wakelam, M.J.O., Murphy, G.J., Hruby, V.J. & Houslay, M.D. *Nature* **323**, 68–70 (1986).
2. Petersen, O.H. & Bear, C. *Nature* **323**, 18 (1986).
3. Farese, R.V. *Endocr. Rev.* **4**, 78–95 (1983).
4. Farese, R.V. *et al. Biochem. biophys. Res. Commun.* **135**, 742–748 (1986).
5. Whitley, G. St. J., Bell, J.B.C., Tait, J.F. & Tait, S.A.S. *Proc. R. Soc. B* **223**, 273–294 (1984).
6. Tait, J.F., Chu, F.W., Hyatt, P.J., Tait, S.A.S. & Whitley, G. St. J. *Sperone Symp. Corticosteroids and Peptide Hormones in Hypertension* (in the press).
7. Whitley, G. St. J., Hyatt, P.J. & Tait, J.F. *Steroids* (submitted).
8. Schwyzer, R. A. *Rev. Biochem.* **33**, 259–286 (1964).

Achieved spacetime infinity

SIR—In his review¹ of my book with John Barrow, *The Anthropic Cosmological Principle*, Press accuses us of indulging in "a distressing amount of mathematical flim-flam . . . For example, . . . Barrow and Tipler [claim] 'A Penrose diagram allows us to define rigorously "an achieved infinity", a concept whose logical consistency philosophers have been doubtful about for thousands of years'. This is a silly assertion, but it is put forth with the utmost gravity, in such a way that many readers will be taken in. And it is only one of many such cases." I cannot comment on the "many such cases" which

are not given in the review, but the "silly assertion" is, perhaps, a case of misunderstanding. Philosophers have debated whether the notion of 'an achieved infinity' is a meaningful concept since Aristotle questioned the notion in his *Physics*; see Sorabji² for a discussion and a history of the debate up to the Middle Ages. The debate is still current; Popper and Wittgenstein³ had a sharp exchange on the question in 1946.

As one will see in the context of the book, we meant by an 'achieved infinity' a spacetime singularity. The Penrose diagram, or more generally the c-boundary construction, allows us to define a spacetime singularity in a mathematically precise way⁴ by attaching a boundary to spacetime. The c-boundary construction is very elegant and natural for globally hyperbolic spacetimes, the only class of spacetimes we considered. In the c-boundary construction, a regular spacetime point is identified with the past light cone $I^-(\gamma)$ of a future-directed time-like curve γ that terminates in that point, and the future c-boundary points are identified with the past light cones $I^-(\gamma)$ of future-directed time-like curves that have no future endpoints in the spacetime. (Past c-boundary points are defined analogously using future light cones.) If a time-like curve defining the c-boundary point is incomplete (that is, of finite length) the point is said⁵ to be a 'singular c-boundary point'. These future singular c-boundary points are the future singularities.

It is generally accepted among relativists that if in fact the Universe is closed and terminates in a final singularity in finite proper time, then the Universe will actually reach this singularity. Strictly speaking, this is not true, since the c-boundary points are not in the spacetime, but there seems to be a general consensus to stretch the meaning of the word 'reach' in this case. Thus the final singularity really exists if the model of gravitational collapse is accurate. From the point of view of the c-boundary construction, the ontological status of the regular spacetime points and the future c-boundary points is the same: they are both the past light cones of time-like curves. Thus if one set of points really exist, then so does the other, and we agree that spacetime really exists. Furthermore, if the initial c-boundary of the Universe is an initial singularity, we can receive information from this singularity just as we can receive information from any other regular spacetime point in our past light cone. Thus it seems reasonable to say it really exists, if such is the initial c-boundary.

The singular c-boundary points are reasonably thought of as infinite because they are (in general) points at which physical quantities are actually infinite. Thus spacetime singularities — if they are indeed on the c-boundary of the actual Universe — are an achieved infinity and are

precisely defined by the Penrose construction. I would have thought that this assertion is the overwhelming consensus of the relativity community.

FRANK J. TIPLER

*Departments of Mathematics and Physics,
Tulane University, New Orleans,
Louisiana 70118, USA*

1. Press, W.H. *Nature* **320**, 315–316 (1986).
2. Sorabji, R. *Time, Creation, and the Continuum* Ch.14 (Duckworth, London 1983).
3. Popper, K. *Unended Quest: An Intellectual Autobiography*, 15–16; 123 (Fontana, London, 1976).
4. Hawking, S.W. & Ellis, G.F.R. *The Large Scale Structure of Space-Time* (Cambridge University Press, 1973).
5. Penrose, R. in *Theoretical Principles in Astrophysics and Relativity* (eds Lebovitz, N.R., Reid, W.H. & Vandervoort, P.O.) 217–243 (University of Chicago Press, 1978).

Estimating guild sizes

SIR—John Lawton¹ has drawn attention in *News and Views* to the important simulation model of Shorrocks and Rosewell² which assumes that aggregation of individual *Drosophila* species is reflected through the parameter k of the negative binomial distribution. The model is based on successive sums of such distributions for two species. The resulting summed distribution is assumed to be negative binomial also, with parameter k_s , equal to the sum $k_1 + k_2$ of those of the component species.

Shorrocks and Rosewell correctly note that this is precisely true only if $\mu_1 k_2 = \mu_2 k_1$, but claim that with equal mean values of 10 or more and values of $k < 2$, it is a close approximation. This claim is technically wrong. Assuming the summed distribution is approximately negative binomial, and that the component species have equal mean values, the method of moments estimator of k_s may easily be found to be

$$\hat{k}_s = \frac{4k_1 k_2}{k_1 + k_2}$$

The approximation does not depend upon the mean value, and the mean value does not need to be 10 or more.

More importantly, however, unless $k_1 + k_2$ is close to k_s , the approximation will be poor. To see the scale of error involved take two examples: first, $k_1 = 7/4$, $k_2 = 1/4$ gives a claimed k_s of 2 compared with estimated k_s of 7/8; second, $k_1 = 2$, $k_2 = 1/2$ gives a claimed k_s of 5/2 compared with estimated k_s of 8/5. Both examples show the approximation is not at all close. The effect that this and the assumption of equal mean values will have on the model requires thorough investigation, although it appears the correction would generate larger estimates of guild size than those estimated by Shorrocks and Rosewell which, as the authors themselves noted, were not quite as large as values from their field studies.

JOE PERRY

*Rothamsted Experimental Station
Harpenden, Hertfordshire AL5 2JQ, UK*

1. Lawton, J. *Nature* **323**, 398 (1986).
2. Shorrocks, B. & Rosewell, J. *J. anim. Ecol.* **55**, 527 (1986).

Aluminium leaching from cooking utensils

SIR—Aluminium is known to have neurotoxic effects^{1–3}. With the discovery that abnormally high levels of aluminium are present in senile plaques in Alzheimer's dementia^{3–5}, the cumulative effects of aluminium poisoning and the question of how this metal enters the body become problems that need immediate attention. Recently, Coriat and Gillard have drawn attention to the high aluminium content of tea leaves and their infusions; aluminium compounds can also be found in water and can be released from utensils during cooking^{3–7}. We have found that the leaching of aluminium from utensils is dramatically enhanced in the presence of trace quantities of fluoride ion.

In an experiment conducted to estimate the rate of leaching, we have found that the presence of only 1 p.p.m. of fluoride (the permitted level of fluoridation⁸) in water adjusted with citric acid or sodium bicarbonate to pH ~ 3 (a pH often realized in cooking conditions) and boiled in an aluminium vessel, liberates nearly 200 p.p.m. of aluminium in 10 min, compared with less than 0.2 p.p.m. in the absence of fluoride. Prolonged boiling produces a concentration of ~ 600 p.p.m., which is reached more quickly the larger the surface-to-volume ratio of the water. The rate of dissolution is pH-dependent with a minimum at neutral pH. Crushed tomatoes (50g in 250 ml) cooked in the same vessel with 1 p.p.m. of fluoride produced a concentration of ~ 150 p.p.m. of aluminium in 10 minutes. Water consumed in some localities contains 10 p.p.m. or more of fluoride. And cooking or prolonged storage in aluminium ware of foods that contain large amounts of fluoride⁹ (~ 500 p.p.m. in tea, 100–700 p.p.m. in fish) could easily release more than 100 p.p.m. of aluminium.

The normal resistance of aluminium to corrosion in mildly acidic or alkaline solutions depends on the formation of an inert oxide film. The corrosion in the presence of fluoride perhaps results from a permeability of the oxide film to fluoride ions, which disrupt the protective film.

K. TENNAKONE

S. WICKRAMANAYAKE

*Institute of Fundamental Studies,
Hantana, Kandy, Sri Lanka
and Department of Physics,
University of Ruhuna, Matara, Sri Lanka*

1. Wils, M.R. & Savory, J. *Lancet* **ii**, 27–35 (1983).
2. Brusewitz, S. *Aluminium*, University of Stockholm Institute of Physics Report 18–11 (1984).
3. Wurtman, R.J. *Sci. Am.* **252**, 48–56 (1986).
4. Duckett, S. & Galle, P. *C. r. hebdom. Séanc. Acad. Sci. Paris* **282**, 393–395 (1976).
5. Candy, J.M. *et al. Lancet* **i**, 354–357 (1986).
6. Coriat, A.M. & Gillard, R.D. *Nature* **321**, 570 (1986).
7. Batchelor, B. *et al. Envir. Sci. Technol.* **20**, 891–894 (1986).
8. Pike, R.L. & Brown, M.L. *Nutrition: An Integrated Approach* 3rd edn, 190–191 (Wiley, New York 1984).
9. Aswathnarayana, U. *et al. Proc. Int. Symp. Geochem. Health* (Royal Society, London, 1985).

A merchant of light

R. V. Jones

The Griffin: The Greatest Untold Espionage Story of World War II.

By Arnold Kramish. Houghton Mifflin: 1986. Pp. 294. \$17.95.

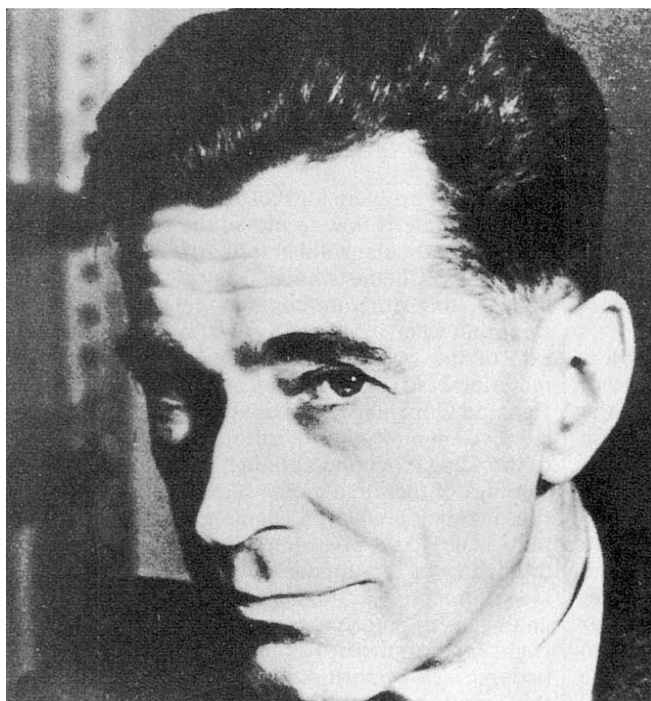
"THE GRIFFIN" of this book's title was Paul Rosbaud, who, as science adviser to the great German publishing house of Springer in the years before 1939, was acquainted with many scientists throughout Western Europe. He travelled widely and frequently, both in search of authors and in promoting the firm's publications.

Rosbaud, who had been born in 1896 and brought up as a Catholic, served with distinction in the Austrian Army in the First World War, after which he became a physics student in Berlin and took a doctorate at the Technische Hochschule. He then moved into publishing, joining Springer in the early 1930s. Alarmed by the rise of the Nazis, concerned by their treatment of his Jewish friends and holding a deep faith in the international character both of science and of the humanities, he used his many contacts to do all that he could to ease the problems of those whose welfare — and indeed lives — were increasingly under threat. And not only those threatened by the Nazis, for in 1935 he also tried to persuade the Russians to let Kapitza return to Cambridge after they had trapped him on a visit to Leningrad.

With so many contacts in the world of pre-war science and with frequent opportunities for travel, Rosbaud was a natural courier of messages between scientific friends across national frontiers; and, from 1939 onwards, some of these messages could provide a means of intelligence. Unfortunately, he had known and liked a member of the British Intelligence Service, Frank Foley, for some years. Because he remained in Germany, he could maintain a fringe contact with German research laboratories, and from time to time he succeeded in getting information passed to London.

Rosbaud's life, particularly as an active agent in wartime Germany with ostensible reasons for visits abroad, is thus a promising subject for a biographer. It has now found one in Arnold Kramish, who was at one time a physicist in the American Atomic Energy Commission and at others attached to the Rand Corporation and the London School of Economics. Kramish

has assiduously gathered details of Rosbaud's life and has delved into the murky world of Intelligence with considerable success. He has discovered much about Rosbaud that I for one did not know, even though I saw the most crucial of Rosbaud's reports that were successfully transmitted during the war, and though I



Paul Rosbaud — "a natural courier of messages between scientific friends across national frontiers;..."

came to know him fairly well afterwards when he lived in London and he wanted me to be the British editor of the new *Handbuch der Physik* — he had returned to publishing, only to have an unhappy relationship with Robert Maxwell who was, I believe, indebted to him for the name "Pergamon".

The clearest of Rosbaud's wartime contributions for which I can personally vouch was his letting us know about the doings and movements of the German nuclear physicists headed by Werner Heisenberg. Rosbaud was intensely alive to the possible release of nuclear energy, for it was he who had rushed the paper by Hahn and Strassmann on the new elements created by the neutron bombardment of uranium into the columns of *Naturwissenschaften* in early 1939. Rosbaud's wartime reports were particularly valuable because they led us correctly to conclude that work in Germany towards

the release of nuclear energy had at no time gone beyond the research stage: his information thus calmed the fears that might otherwise have beset us. Kramish's description of Rosbaud's dangerous work in this connection by itself justifies the biography.

I cannot speak with such confidence about Rosbaud's other contributions to scientific intelligence that Kramish describes. Where he discusses events and characters with which I had direct contact, I sometimes have difficulty in reconciling his account with my own memories. He is handicapped by having to speculate too far from an inevitable sparsity of known facts, particularly on events inside the

British Intelligence Service. Some of his speculations seem tendentious, and some of his descriptions are quite erroneous — "Sir Edward Appleton of the S.I.S.", for example, at least if S.I.S. has its normal connotation of the Secret Intelligence Service. Rutherford is surprisingly described as "certainly a theoretician" in contrast with Cockcroft; the V1 flying bomb was not a glider; and, so far as I know, phosphorescent materials are not "an important ingredient of heat-seeking sensors that guide missiles to their targets". Such examples contrast with the punctilious accuracy that Kramish sometimes achieves.

In other places he appears to have aspired to verisimilitude by adding naturalistic detail that he could not possibly know — a charge that he himself properly lays in one instance against another author, Anthony Cave Brown. This concerns the mode of delivery of the Oslo Report,

the package of information on new German weapons and techniques that was sent to the British Legation in Oslo from an anonymous source early in November 1939. The Report told us, for example, of a new radio ranging system that we later found the German pathfinders using in the Blitz, and of the two principal types of German radar equipment which we later discovered under the code-names Freya and Würzburg; other items included acoustically homing torpedoes, radio-controlled glider bombs, and large rockets with gyroscopic control. Peenemünde was also mentioned. Kramish himself claims to have discovered "beyond doubt" that the Report came from Rosbaud, and that "it seems indisputable" that it was delivered to the Legation by the hand of Odd Hassel, who later won a Nobel Prize.

It can be argued convincingly that Rosbaud had the qualifications and the motive to write the Report. He could perhaps also

have had the knowledge. Kramish speculates that Rosbaud had the opportunity to visit Oslo in November 1939; but he has to admit that although Rosbaud was there in September, Rosbaud himself stated after the war that he had no further contact with Hassel until December. So Kramish has to add to the conjectures with which the book is peppered the further one that Rosbaud's memory was at fault.

As for the speculation that Hassel handed the Report into the Legation, Kramish appears to have overlooked the positive statement by the officer who received it, the Naval Attaché, Captain Boyes, that it came in the form of letters "posted in Norway". That statement was available to Kramish, although he does not mention it. Had he done so, he might have proposed

that Captain Boyes's memory, like Rosbaud's, was at fault.

Actually there is no need for either speculation, for Rosbaud did not write the Oslo Report: to my positive knowledge he had nothing to do either with its provenance or its transmission. At the same time, despite this substantial error, Kramish has performed a welcome service in ensuring a wider appreciation of those genuine and important contributions that Rosbaud so courageously made. □

R.V. Jones, 8 Queen's Terrace, Aberdeen AB1 1XL, UK, is Emeritus Professor of Natural Philosophy in the University of Aberdeen. He was the member of Air Staff responsible for scientific intelligence during the Second World War, and is author of Most Secret War (Hamish Hamilton, 1978), which describes the activities of the scientific intelligence services during the war.

Brief encounters

Jeremy K. Burdett

New Directions in Solid State Chemistry. By C.N.R. Rao and J. Gopalakrishnan. Cambridge University Press: 1986. Pp. 516. £55, \$79.50.

SOLID state chemistry is a tough subject and for many years was rather unfashionable. Synthetic techniques are at the Neanderthal stage compared to those used in organic synthesis. Characterization of the reaction product is invariably difficult, because it is often non-crystalline — in these cases, X-ray crystallography, a major tool for the molecular inorganic and organometallic chemist, is inapplicable, and NMR, a mainstay of molecular organic chemistry, is neither as easy to perform nor as easy to interpret. The materials themselves are frequently non-stoichiometric and often the presence of a tiny amount of impurity can dramatically change the crystal structure. In addition, the description of solid state structures, relating the collection of *x*, *y*, *z* coordinates to some geometrical model in terms of linkages between atoms in close contact, sometimes presents quite a challenge. All of these points emerge from a close reading of Rao and Gopalakrishnan's new book, in which the authors attempt to paint a broad picture of our current knowledge of solid state chemistry. At the moment, the field is experiencing a resurgence of interest from chemists of all persuasions, and the appearance of a work of this type is clearly timely.

New Directions, however, is written as a very long review article. It has the flavour of a Chemical Society *Specialist Periodic Report*, in that a large number of topics are covered in each chapter, many of which are described in a sentence or two and then labelled with a reference. The result is that the book provides only a

sketchy overview of the area as a whole, and gives little new insight into some of the tantalizing problems mentioned above.

A further problem for those working on the solid state is how to model the structure of a particular solid in terms of chemical bonds. Chemists have always leaned heavily on structure/bonding relationships, but when it comes to solids, description of the way some of these complex materials hold together is not at all simple. Rao and Gopalakrishnan use the traditional division into solids of different chemical bonding type, and although the shortcomings of the ionic model for coordination number prediction (Pauling's radius ratio rule) are presented, no incisive discussion of this fundamental topic follows. Similarly, although the solid state analogue of molecular orbital theory (tight-binding theory) is mentioned, the pedagogical opportunity presented by this correspondence is not pursued.

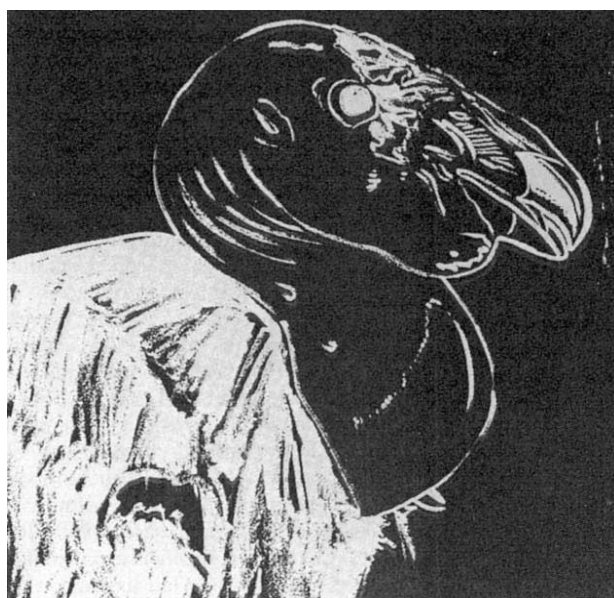
Although it is entitled *New Directions*,

the book is somewhat backward looking. Much of the material describes recent work using traditional concepts and philosophy. For example, studies of the organic solid state in research on conducting polymers have led to some remarkable collaborations between physicists and chemists, perhaps one of the success stories in this interdisciplinary area. A startling variety of species of semiconducting, superconducting and metallic solids have been produced, the properties of which may often be subtly adjusted by changing pressure or the nature of an apparently innocent atom or grouping in the organic building blocks. Polyacetylene and (SN)_x are now textbook examples of these types of materials, and yet less than 12 lines are used to describe them and the science behind them.

Another area which does not get its due is the electronic structure of solids. Large strides have been made in recent years in understanding the crystal structures of extended arrays, but only passing reference is made to the work of the new breed of physicists, chemists and chemical physicists who are intent on linking structure and properties of materials. For example, it has been possible for at least half-a-dozen years to predict by calculation from a list of possibilities the lowest energy structural alternative (admittedly, for simple systems only).

Altogether, the book would serve quite well as a relatively brief introduction to those new to solid state chemistry. But the authors have missed the opportunity to present some of the novel developments of the past few years and convey the excitement which is now percolating through this area. □

Jeremy K. Burdett is a Professor in The James Franck Institute and the Department of Chemistry, University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, USA.



Detail from a painting by Andy Warhol of the California condor; the original is in yellow and red on black. The picture is reproduced from Vanishing Animals, an account in words (by Kurt Benirschke) and pictures (by Mr Warhol) of the plight of 15 species of endangered animal. Publisher is Springer-Verlag, price is DM 120, \$49.50.

Culture in danger

Terence R. Lee

Risk Acceptability According to the Social Sciences. By Mary Douglas. *Russell Sage Foundation: 1986. Pp. 115. Distributed by Basic Books, New York, pbk \$6.95.*

Risk assessment and the concept of acceptable risk are much in vogue of late, and there are few learned scientific societies that have not felt the need to confer on the subject. But this book is not in the mainstream of such considerations and does not mention probabilistic risk assessment, fatal accident frequency rates or LD₅₀ tests. Rather, it is about risk *perceptions* as products of culture.

In the past, approaches to risk through the social sciences have been largely dominated by psychometric studies of perceived risk and by classical utility theory with its later refinements. Psychologists, economists and others in the field are acutely aware of the vast discrepancies between the public's idea of risks and the scientific estimates of the risk assessor. But they decline to attribute them to public ignorance or "irrationality".

Workers in comparative risk perception have emphasized the profile of qualities that go to make up the "meaning" that a hazardous activity has for people. They demonstrate, empirically, that (for example) voluntariness, and familiarity with the risk, its catastrophic potential, its threat to self or to society and the immediacy of its effects are all potent intervening variables. In utility theory, the thrust is towards the complex trade-offs that people make between costs (including risks) and benefits, together with, for example, the "concept of bounded rationality" which explains the changes in the trade-off when the bottom line has to be preserved.

Both of these approaches and more besides are admirably reviewed in Mary Douglas's book. Their limitations are exposed ruthlessly but politely. The author is neither a psychologist nor an economist (despite her erudition in these fields) but an anthropologist. She argues that "...what is needed is to persuade the two sets of analysts that there is a mediating element, *culture*, which is worthy of their combined attention". This is undeniable, though one may be forgiven for mentioning that the book's somewhat single-minded proselytization of an anthropological perspective may itself demonstrate the force of its central argument — there are no stronger cultural influences on the perception of risk than disciplinary affiliations!

In defence of psychology it has to be said that, although it has shamefully neglected the importance of social factors, it is

far from unaware of them. It is a parody to suppose that psychologists expect to identify explanatory variables that are determined genetically or in early childhood, and thereafter become part of an enduring personality. What they do seek, however, are broad generalizations, and they certainly recognize that the qualities of familiarity and voluntariness, for example, are culturally determined. Further, the systematic differences in perception between members of different cultural groups (namely the League of Women Voters, upwardly mobile businessmen, students and "experts") were demonstrated by psychologists some years ago. Much of personality is internalized culture.

Once again, a great deal depends on the most profitable level of analysis; there is a

sense in which it helps very little to argue that racing drivers, driving-school instructors and housewives all have different perceptions of traffic risks because they are conditioned by different cultural backgrounds. The answer must surely lie in the classification of cultures in terms of the constructions they put upon different hazards — and why. Mary Douglas has made an excellent start in this direction. It has to be added, though, that the exercise will have to "go empirical" if it is to make progress beyond this set of signposts which at present do little more than point clearly and convincingly in the right direction. □

Terence R. Lee is a Professor in the Department of Psychology, University of Surrey, Guildford GU2 5XH, UK.

Crusading appeal

Alexandra Dixon & Brian Bertram

Animal Extinctions: What Everyone Should Know. Edited by R.J. Hoage. *Smithsonian Institution Press/Eurospan, London: 1986. Pp.192. Pbk \$10.95, £9.45.* **The Last Extinction.** Edited by Les Kaufman and Kenneth Mallory. *MIT Press: 1986. Pp.208. \$16.95, £12.95.*

THE conservation ethic can be likened to a new religion — not in being based on faith, but in the vigour with which those who are convinced of its necessity strive to persuade others to join them. The target is the unconverted general public, the many uninformed, uninterested or unsocial individuals who persist in contributing directly and indirectly to the destruction of the Earth's natural resources, particularly its living resources. Unlike the religious missionary, conservationists are able to marshal a formidable array of scientific and economic arguments to support the cause. The need is for these arguments to be put across to the general public, and urgently.

These two collections of symposium and lecture papers aim to meet this need. The first, *Animal Extinctions*, will arm the already converted. It contains an assemblage of information on the theory and application of conservation biology, covering such topics as strategies for preserving species in the wild, genetics, sustainable exploitation of wildlife and the economic value of species to mankind. Conservation biologists will have heard much of it before, partly because the symposium was held as long ago as 1982, and the material was not new then. The general public will find its presentation dull and the arguments rather heavy going, but it is useful to have this authoritative textbook available at last.

For the unconverted, *The Last Extinc-*

tion is better directed and more effectively evangelical. The editors state in their first paragraph that their aim is to persuade people to rearrange their priorities to avert the threat of mass extinctions, and the contributors set about doing precisely that. The eight public lectures given in Boston, Massachusetts, in 1984 were delivered by a good mixture of representatives from government agencies and the academic, zoo, aquarium and botanical communities in the United States. Using a variety of natural habitats and species to illustrate their points, the authors collectively argue that the looming threat of mass extinctions is *not* a natural event, that it is of fundamental concern to us all and that urgent action is required, to which end they provide clear recommendations and a useful directory of conservation organizations.

While preaching the same overall message, each chapter deals with a distinct issue. "Why the Ark is Sinking" discusses the crucial difference between natural and unnatural extinctions, and why we should care. A chapter on the Amazon illustrates the nitty-gritty realities of local poverty, conflict of interests and ill-considered development, and gives the reader a real idea of the factors involved in conservation on the ground. But conservation is not only about saving exotic species in far-away lands. For instance, few North Americans realize that at least 25 species and sub-species of vertebrate have become extinct in their continent since 1950.

The Last Extinction is compelling reading. It is well written, and attractively presented with a glossy cover and lots of pictures. Absorbing the arguments presented by the authors is quite painless, which is why the book is so effective. The man in the street will enjoy it; he will also learn a lot. He may also join the crusade. We must all hope so. □

Alexandra Dixon and Brian Bertram are at The Zoological Society of London, Regents Park, London NW1 4RY, UK.

Problems of scale in ecology

John Lawton

A Hierarchical Concept of Ecosystems. By R. V. O'Neill, D. L. DeAngelis, J. B. Waide and T. F. H. Allen. *Princeton University Press*: 1986. Pp. 253. Hbk \$45, £29.70; pbk \$14.50, £9.60.

ECOLOGISTS have been slow to wrestle with the importance of spatial and temporal scales in population, community and ecosystem processes. By way of illustration, imagine a coral reef, and the myriads of jewel-like fish that call it home. How important are interspecific interactions between the fish in determining "community structure" — which species live where, when and in what abundance?

Study a single patch of coral a few metres square for a few weeks, and fish may appear to come and go at random: evidence of pattern or structure is hard to find because the spatial scale of observations is too small. Fish communities may, instead, be structured over much larger scales, set by patterns of larval recruitment from the plankton over several kilometres of reef. Or, with a different question in mind, it may be very difficult indeed to link detailed behavioural studies of interactions between fish, over time-scales of minutes or hours, to population dynamic processes taking months or years. Finally, the scuba-diver interested in fish behaviour takes the coral for granted as part of the habitat. Yet corals build and die over decades: how are fish behaviour (involving the fastest time- and the smallest spatial-scales) and coral dynamics (with long time- and huge spatial-scales) to be combined into realistic models of the reef ecosystem?

A Hierarchical Concept of Ecosystems tries to address such questions: they may seem obvious, but they have received far too little attention from ecologists. More often than not, field workers arbitrarily select a scale of observations that "seems right", and theoreticians are mute about space and time. But as O'Neill and his colleagues show, these are risky strategies. Theories and observations appropriate for one kind of ecology do not easily translate to larger and longer or smaller and faster spatial and temporal scales. Instead, the natural world can profitably and more correctly be viewed as a multi-layered system, hierarchical in space and time. One virtue of this approach is that a number of apparent conflicts are immediately resolved. For example, assemblages of species (communities) may indeed be regarded as more or less separate entities, each with its own rich biology, much as H. A. Gleason suggested over 50 years ago,

whilst entire ecosystems may function holistically, fixing energy and cycling nitrogen irrespective of the species involved. Neither point of view is more "correct" than the other. Similarly, populations and communities of organisms may well be equilibrium systems, with population sizes regulated by food or space, embedded in an ecosystem undergoing long-term, very slow changes far from equilibrium. Again the "correct" picture depends upon your perspective and what questions you ask.

The general message of this book is therefore valuable and well put. Its strength lies in the exposition of hierarchy as a way of thinking, in repeatedly drawing attention to the pitfalls of ignoring scale, and in reinterpreting existing data, disagreements and conflicts in the light of the new perspective. In contrast, I found the development of a formal theory of hierarchy, useful for ecologists, more disappointing. The ideas are more qualitative than quantitative, and starting to accumulate their own jargon: it remains to be seen whether ecologists will find the term "holon" useful, and whether holons

can be identified, measured and studied in any objective way. Nor do many clear and new predictions emerge from the book. There are some, for example specific predictions about patterns of nutrient loss from stressed laboratory ecosystems ("microcosms"), and much more general and important ideas about overall control of the dynamics of populations, or other ecosystem components, interacting in nested hierarchies of slow and fast time-scales. I was disappointed not to find more, but recognize that at this stage in the game more rigorous theory and predictions are hard to come by.

The construction of theoretical models of ecological processes as spatial and temporal hierarchies, and attempts to match, test and refine the models against the real world, will undoubtedly be one of the growth industries in ecology over the next decade. *A Hierarchical Concept of Ecosystems* is a good point of departure for this venture, which could prove to be an exciting and unpredictable journey. □

John Lawton is a Professor in the Department of Biology, University of York, Heslington, York YO1 5DD, UK.

Even cognitions

Nicholas J. Wade

Odd Perceptions. By Richard L. Gregory. *Methuen, London and New York*: 1986. Pp. 230. £16.95, \$19.95.

EDITORS of journals usually confine their published remarks to matters of policy. Not so Richard Gregory in his editorials in the journal *Perception*, which have appeared over the past 15 years. These essays reflect the ideas that were exciting him as the deadline for publication descended, and a selection of them has now been collected together under the title *Odd Perceptions*.

Many of the pieces have been amended, extended or amalgamated, and the variety of subject matter is vast: from Bacon to exploratories, from eclipses to ESP, from gnomons to von Neumann, from wind-up frogs to Wittgenstein. They are grouped under three headings — amusing, musing and using (alas, there is no fourth musical section!).

Oddly enough, it is not perception that binds the essays together, but the history of science and its associated methods. At times, the history is confined to individual achievements (such as Kepler's law of planetary motion or Turing's approach to machine intelligence), at others it is addressed to a particular problem (mirror reflections or consciousness, for example). All of these issues tend to be reduced to problems of cognition: how ideas about physical processes develop,

and how cognitive processes can be modelled.

According to Gregory, these two types of problem are closely related: his thesis, presented in more detail in *Mind in Science* (Weidenfeld & Nicolson, 1981), is that theories of mental processes draw upon the current solutions to technical problems. The methods that work with physical systems should, he considers, be applied to biological ones. Again, oddly enough, very little is said about machine vision, even though the approach he adumbrates has had some of its greatest successes in image interpretation. Gregory's cognitive approach to perception (as outlined in his chapter on illusions) was developed without the benefit of these solutions, and so it would have been instructive to learn whether his theory has been revised accordingly. Perception is conceived of as a process generating and testing hypotheses — much as scientists do with respect to physical phenomena — until a resolution of the problem is found. How the hypotheses are generated is one of the many deep mysteries we are left with; indeed, one of the pleasures of this book is that it does not set out to provide the answers, but to frame appropriate questions, and to place them in historical perspective.

The essays are served with a liberal mixture of autobiographical anecdote, wit and word-play. They stand as a tribute to Richard Gregory's enthusiasm for perception — the process and the journal. □

Nicholas J. Wade is a Reader in the Department of Psychology, University of Dundee, Dundee DD1 4HN, UK.

The integrity of the scientific literature

Walter W. Stewart and Ned Feder

A case of admitted scientific fraud has shed new light on the system that ensures the integrity of the scientific literature. Certain lapses from generally accepted standards of research may be more frequent than is commonly believed.

WE wish to report the results of an unusual investigation. In May 1981, the colleagues of a respected young scientist were shocked to discover that he was a forger of data. Initially, Dr John Darsee confessed to a single forgery, but it was subsequently found by his colleagues and by three investigating committees that he had fabricated much of the data that formed the basis of his more than 100 publications¹⁻¹²⁹ over a period of about three years.

Although many of these publications embody data now acknowledged to be fraudulent, we have scrutinized these publications for internal consistency, accuracy and completeness. Our sample consists of the 109 publications by Darsee with the 47 scientists who were his co-authors, all of whom worked at two leading US medical schools. We also studied the reports of the committees set up to investigate the Darsee affair. What is special about the publications in our sample is that, as a by-product of the extensive investigations into Darsee's work, the circumstances surrounding their publication were known in unusual detail.

Our objective was to discover whether the published scientific reports would throw light on the vigilance of referees, of editors of journals and of Darsee's co-authors in meeting the standards conventionally accepted as necessary in the scientific literature. We have found many of the reports to be flawed in ways that could have been recognized by those concerned with their publication.

We emphasize the distinction between the original forgery and the lapses from standard publication practices we have found in our "sample of convenience"¹³⁰. Other studies will be needed to determine whether or not our results are representative of clinical scientists, of biomedical scientists or of scientists in general.

Background

The circumstances of the Darsee affair have been widely reported¹³¹⁻¹⁴¹. In 1978 to 1981 inclusive, he was author or co-author of 18 full-length research papers published in major biomedical journals and of about 100 abstracts, book chapters, letters, reviews and short papers in clinical and experimental cardiology¹⁻¹²⁹. In part because he was so prolific, he was highly regarded

first at Emory University School of Medicine and then at Harvard Medical School.

Five months after Darsee's confession, in May 1981, to a single act of data fabrication, it became apparent that there had been more than one such episode. Ultimately, three committees were appointed to investigate. Harvard appointed a committee headed by R.S. Ross, which issued its report¹³¹ in January 1982. The NHLBI (National Heart, Lung, and Blood Institute) appointed a panel headed by H.E. Morgan to investigate the circumstances at Harvard, where most of the research had been supported by the National Institutes of Health (NIH); this report¹³⁴ was completed in June 1982. Finally, at Emory, an investigating committee headed by N.C. Moran was formed, and issued its report^{132,133} in March 1983.

Shortly after the appearance of the report of the Moran committee, and almost two years after the frauds themselves had come to light, we embarked on our study. We were especially careful to discriminate between the direct effects of Darsee's forgeries, which are not the subject of this report, and the defects in the papers of which he was a co-author. We recognize that some may disagree with some of our decisions as to what constitutes a defect, and with our judgements of the degree to which Darsee's co-authors may be held to have been responsible.

Darsee's publications were in experimental and clinical cardiology, a field of research quite different from our own. For this reason, we have no comments to make on the pointedness or importance of the research described in the publications or about the design of the experiments. No doubt we may also have overlooked some errors in the publications because of our lack of first-hand knowledge of the field. At the same time, we believe that our position as outsiders had the advantage that none of those who published with Darsee was known to us personally or professionally. It may also be an advantage that we have been able to carry out our analysis in the same way that other scientists outside the field of cardiology might approach such a task.

Of Darsee's 47 co-authors, 24 worked at Emory University School of Medicine and 23 at Harvard Medical School. Biographi-

cal data were found for 44 of the 47 co-authors. Thirty-nine were apparently MDs, one was a PhD and one held both qualifications; three apparently had no advanced degree. At the time of their involvement with Darsee, about half of the co-authors were 30-40 years old and had received their advanced degrees 5-15 years previously. Approximately equal numbers were senior and junior to this group, the latter including technicians, medical students, resident staff at teaching hospitals and junior faculty members. The senior co-authors included professors and department chairmen. (Darsee himself was 33 in 1981.)

Abundance of errors

Many of the published errors and discrepancies were presumably introduced by Darsee. We nonetheless considered all errors, major and minor, to be of interest because they might reflect on the care with which the co-authors had checked the publications that bore their names, and might reflect indirectly on the care with which they had carried out their share of the research.

All but a few of the 18 research papers^{6,24,26,50,53,87,96} (Fig. 1) and many of the abstracts contain errors or discrepancies that can be recognized simply by examining them carefully. We emphasize again that many of the errors and discrepancies are minor, for example, a small discrepancy between a numerical value printed in a table and what is supposed to be the same value printed in the text. Some errors,

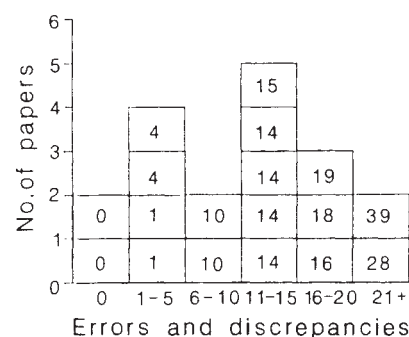


Fig. 1 Errors and discrepancies in published papers. Each box represents a single paper, and the number within the box is the sum of the errors and discrepancies for that paper.

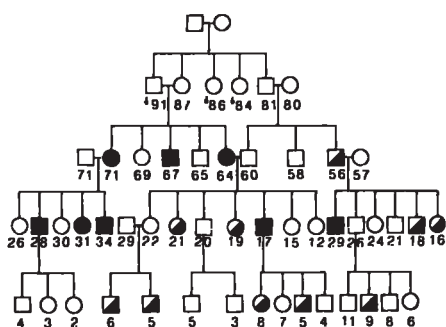


Fig. 2 Pedigree reproduced from ref. 87. The age of each family member at the start of the investigation is given below each symbol. Arrows (second row) denote those who had died before the start of the investigation. Note the 17-year-old male (row 4) with children ages 8, 7, 5 and 4; at the age of 8 or 9 (probably 8) he impregnated the mother of his oldest child.

however, are so glaring as to offend common sense.

We present a detailed analysis of three sample research papers from which the reader may judge the types of errors we counted. Two papers^{87,88} from Emory and one from Harvard⁹⁶ are discussed below as examples of our procedure. Two of these^{87,96} were eventually retracted wholly^{127,142} or in part¹⁴³, but one⁸⁸ has not been retracted and is still cited as valid^{132,133}.

The most striking example, in ref. 87, is that in which the authors reported an arresting pedigree depicting a family with a high incidence of an unusual form of heart disease. Inspection of the pedigree (Fig. 2), reveals that a 17-year-old male had 4 children, ages 8, 7, 5 and 4. (It is unlikely that the father's age is a misprint: it appears in the figure and in two places in the text.) Thus the father was 8 or 9 (probably 8) when he impregnated the mother of his first child and 9 or 10 when he impregnated the mother of his second. This bizarre feature of the pedigree, which perhaps could have raised questions about the validity of the entire paper, was apparently not noticed by co-authors or referees, nor were the following unlikely groupings of ages: his sister, brother and first cousin had their first children at ages 16, 15 and 15, and three women in the preceding generation had their last children at ages 41, 45 and 52.

The summary of the paper gives the urinary taurine levels of the family members with congestive cardiomyopathy as ranging from 411 to 536 mg taurine per g creatinine, but the text and Table 1 of the same paper give for the same measurement 426 ± 45 mg taurine per g creatinine (mean $\pm$ standard deviation). Simple inspection suggests that these two sets of numbers cannot simultaneously be valid. Likewise, the summary of the paper states that the urinary taurine levels of the family

members with mitral valve prolapse ranged from 215 to 265 mg taurine per g creatinine, whereas the text and Table 1 give 220 ± 31 mg taurine per g creatinine. These values are central to the thesis of the whole paper. The paper has three numerical discrepancies with ref. 40, and five with refs 19 and 28, abstracts which appeared before the paper and which deal with the same family. All three abstracts of this work indicated that, at the beginning of the study, all 46 members of the kindred were alive, which contradicts Fig. 1 of ref. 87, in which 3 of the 5 oldest are shown as having died before the beginning of the study.

A more subtle example of a major error is illustrated in Fig. 3, taken from a paper⁸⁸ also from Emory by six members of our sample. The figure depicts changes in haemodynamic values in five patients at various time intervals after a surgical procedure. Almost all the data in this paper⁸⁸ are contained in three figures which cannot be reconciled with each other. The mean and standard deviation for haemodynamic values in the five patients are given in our Fig. 3 (Fig. 1 in the original paper). It is easy to see that many error bars are the same size, and also that mean SVR (systemic vascular resistance) is almost constant at -160 units (the bottom of the graph) in four serial measurements. (Abbreviations are explained in the legend to Table 1.)

These results are especially implausible in the light of data given elsewhere in the paper, where it is shown that SVR and other haemodynamic values for two individual patients vary markedly in the course of time. Moreover, many standard deviations in the figure we reproduce appear on inspection too small to be reconciled with the widely scattered values shown elsewhere in the paper; this is true

of SVR at 0.5 h, PVR at 1 h, SWI and SVR at 8 h, and CO at 12 h. Indeed, simple calculations show that it is virtually impossible to reconcile most of the data in the figure we reproduce with the data in the other two figures of the paper.

The introductory section in this paper⁸⁸ indicates that the study of five patients was carried out "in four uncomplicated cases of peritoneovenous shunt insertion, and in one subject who developed signs of cardiac failure and pulmonary edema", but the figure legends indicate that two of the five patients developed pulmonary oedema. Specifically, one of the four patients said to be without complications "developed clinical and radiographic signs of pulmonary edema 45 minutes after the shunt was opened" (Fig. 2 legend). The other patient, whose pulmonary oedema was more severe, is described variously as developing this complication "one hour after the shunt was opened" (Fig. 3 legend) and "within a few hours" after the shunt was opened (results section⁸⁸).

Furthermore, the statements that the changes of the mean shown in Fig. 1 after the first half hour were "very small" and that there was "very little change" appear to be contradicted by the actual maximum changes of the mean — a factor of two or more in the change from the value at time zero (corresponding to roughly a 25–60% change in the mean itself) — for three of the five parameters in the figure.

There are obvious inconsistencies between this paper⁸⁸ and a preceding abstract³⁰ describing the same five patients. Eight of ten values for mean or standard deviation given in the two publications are inconsistent (Table 1). This includes most of the numerical data in the abstract. Another comparison (Table 1) shows discrepancies in two out of five standard deviations between one

Table 1 Discrepancies between a paper and two abstracts

	Ref. 88	Ref. 30	Refs 20 and 88	Refs 20 and 30
Mean or s.d.	m. $\pm$ s.d.	m. $\pm$ s.d.	s.d.	s.d.
Parameter measured:				
CO	2.1 ± 0.8	$2.6^* \pm 0.8$	$1.0^\dagger$	0.9
PCWP	10.0 ± 3.7	$12^* \pm 2^*$	4.4	1.8
PVR	-43 ± 33	$-64^* \pm 24^*$	30	$43^\ddagger$
SVR	-123 ± 33	$-192^* \pm 168^*$	$199^\dagger$	158
SWI	10.6 ± 4.2	$11.3 \pm 5.8^*$	$4.8^\dagger$	$5.2^\ddagger$

Ref. 88 is a full-length paper; refs 20 and 30 are abstracts. Discrepancies are indicated between columns 2 and 1 (*), columns 3 and 1 (†), and columns 4 and 2 (‡). Values in column 1 were obtained by measurements of the 0.5 h points in Fig. 1 of ref. 88. The values in column 2 were copied directly from ref. 30, except that for SVR and PVR, the R units³⁰ have been converted to dyn s cm^{-5} (factor of 80; ref. 144). Columns 3 and 4 contain standard deviations for 5 patients calculated from the means and standard deviations given in ref. 20 for 4 patients and the means given in refs 88 or 30 for 5 patients. The values were derived from calculations based on "population" standard deviations; calculations based on "sample" standard deviations gave similar values and conclusions. When comparing values, allowance was made for rounding of the final digit in refs 20 and 30 and for uncertainties of measurement in ref. 88. Abbreviations: CO, cardiac output; PCWP, pulmonary capillary wedge pressure; PVR, pulmonary vascular resistance; SVR, systemic vascular resistance; SWI, stroke work index. Units are given in the publications, except as noted above.

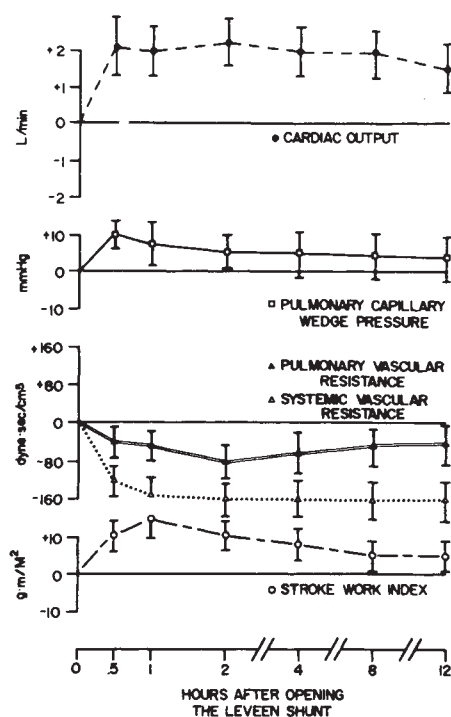


Fig. 3 Reproduction of Fig. 1 from ref. 88. The graph depicts the mean and standard deviation for changes in haemodynamic values of all five patients in the study. Note the remarkable similarity in size of many error bars. Note also the relative constancy of SVR (average value for five patients) at -160 for four serial measurements, despite the variability of SVR (individual values for two patients) in Figs 2 and 3 of ref. 88 (not reproduced here).

abstract²⁰ and the other³⁰, and in three of five standard deviations between the abstract²⁰ and the paper⁸⁸. Thus the two abstracts are inconsistent with each other and with the full-length paper.

In both research papers described above, the major errors are fundamental and

call into question the validity of the paper's main conclusions. More typically, however, the errors simply reflect on the care with which the paper had been prepared.

The final example was contributed to a major journal by one of the co-authors about two months after Darsee had admitted to him and another co-author that he had forged data in another study. At the time the paper⁹⁶ was contributed, these two co-authors were carrying out, privately, a "detailed review and analysis"¹³¹ of the extent of Darsee's forgery in their laboratory. Thus these two scientists were simultaneously publishing a paper containing the results of their collaboration with Darsee and serving in private as investigators into the extent of his forgery of data. One author stated¹³⁸ that he had "total confidence" in this particular paper, which he also said¹³⁵ had "passed special scrutiny". The Ross Committee, asked to review Darsee's work at Harvard and prepare a statement of record, affirmed the co-authors' view that this publication was accurate.

The paper⁹⁶ describes the recovery of heart muscle injured reversibly by a brief experimental occlusion of a coronary artery. One measure of heart muscle function (systolic shortening) is said in the legend to Fig. 3 not to recover fully "until after 7 days of reperfusion". This description is at least unclear, and we regard it as erroneous, since the figure itself and the data in Table 1 show full recovery at or before 7 days.

We present this as an example of an error of very minor significance, but one that we think suggests that the co-authors were not sufficiently careful in examining their own paper before publication. There are small discrepancies between the summary (p. 7152) and the text (p. 7153) with regard to ATP concentration at both 90 min and 72 h of reperfusion. There is a discrepancy between the error bar for ischaemic subendocardium at 15 min of occlusion in Fig. 1A and the corresponding number in the summary and text.

The paper also contains discrepancies with another paper⁹⁵ and with an abstract¹⁰⁵, both published previously, and both describing similar or identical studies on similar groups of dogs. The values of maximum negative dp/dt (all between about $-4,000$ and $-4,300$ mm Hg s^{-1}) for the dog in Fig. 3 of ref. 96 cannot be reconciled with the values (means from $-2,248$ to $-2,562$; standard deviations from 200 to 316) presumably calculated from this dog plus 6 others and listed in Table 2 of ref. 95. Also, the size of the experimental groups for measurement of ATP and creatinine phosphate at 15 min occlusion and 90 min reperfusion is different in the three publications: 14 or 21 dogs (the text is ambiguous) in ref. 95; 34 or 41 in ref. 96; and 36 in ref. 105.

Despite these differences in the numbers of dogs in the experimental groups, the means and standard errors, wherever given, are the same for all three publications. For example, in refs 95 and 96 (Figs 1 and 2, 15 min occlusion and 90 min reperfusion), there are 16 values and 16 error bars; these are identical in the two papers. Also, in the cases where numerical data are given in the text of the three publications, the numbers after rounding are the same. Such a coincidence of means and standard errors from experimental groups of different sizes is extremely implausible. The paper⁹⁶ contains additional errors and discrepancies not described here.

Among the 18 research papers, there were as many as 39 errors and discrepancies in a single research paper, with an average of about 12 per paper (Fig. 1). Of the 22 scientists who were co-authors of a research paper, 19 were co-authors of at least one research paper containing 10 or more errors or discrepancies.

The analysis of errors and discrepancies given above is based entirely on our sample of publications in scientific journals. In contrast, some of the analyses that follow depend on information given in the committee reports.

Retention of data

The reports of the investigating committees showed that in almost all cases most of the co-authors had not retained the experimental or clinical data on which the publications were based. This was true of 7 of the 8 Harvard papers, which involved experiments on dogs, and of 8 of the 10 papers from Emory, 8 of which involved human subjects. [These figures appear mostly attributable to the failure to retain data by Darsee, the chief source of experimental data — Editor, *Nature*.] In the latter case, there were 7 papers for which some of the co-authors had retained neither names nor other identifying information for the human subjects. For the purposes of the present report we have disregarded all papers not involving human subjects except for the cases discussed below.

Of 14 who were co-authors of papers (all from Emory) involving human subjects, at least nine failed to retain the list of subjects (Table 2). One paper²⁵ is particularly interesting: a report on the incidence of mitral valve prolapse in presumably healthy young men. The paper is considered valid¹³³ and is heavily cited¹⁴⁵. The co-authors "maintain that the data in the paper are valid although none of them has a copy of the original research results. They state that they did the study and oversaw the writing of the paper"¹³³. Although Darsee was first author of the paper, the Moran report provides no information on his role in the project and, in particular, on his opportunities for altering or forging the data; possibly this in-

Table 2 Frequency of lapses classified by publication and by co-author

Type of lapse	No. of publications	No. of co-authors
1	28	26
2	42	21
3	7	9
4	13	6
5	7	6
6	1	4
7	8	8
8	1	1
9	1	1
10	2	3

Column 1, type of lapse as defined in Table 5. Column 2, number of publications by co-authors in which there were one or more examples of the lapse in column 1. Column 3, number of co-author involved in each type of lapse. Some publications and some co-authors were associated with more than a single type of lapse.

formation is contained in the confidential appendices to the report. Thus the only available support for the validity of this paper is based on a statement attributed by the Moran Committee to the co-authors, given without supporting data or details.

Involvement in research

In considering involvement, we confined our attention to the 18 research papers. We defined "direct involvement" to mean actual participation in the acquisition and interpretation of data, or the making of an intellectual contribution which was essential to the research. We disregarded a person's rank in the laboratory hierarchy in judging whether an author was directly involved.

Our evidence for the extent of an author's involvement came from the reports of the investigating committees. For example, according to a committee report¹³², one co-author's role was to encourage Darsee and provide some of the grant support for the research. In another committee report¹³⁴ a senior scientist is described as having other duties not allowing him sufficient time for direct participation in the research laboratory and also as being seldom in the laboratory and having no direct supervisory role. In both these cases we considered that the co-authors had not met our criteria for direct involvement, in the sense of direct participation in experimental work. Supervisory functions may be exercised with a greater or lesser degree of remoteness, and we acknowledge that the evidence available is limited and that our judgements on this point are necessarily subjective.

We consider those not directly involved in the research to be "honorary" authors. Honorary authorship was a common occurrence, according to our analysis. Of 18 papers, 13 had at least one author judged by us to be honorary. There was a total of 49 authorships (that is, occurrences of an author's name on one of the 18 research papers); in 16 authorships (33% of the total), involving 6 of the 47 co-authors, the author was judged to have little or no direct involvement (Table 2). Those who were honorary authors at least once also published significantly more papers than those who were never honorary authors (3.33 ± 1.63 versus 0.71 ± 1.42 , mean $\pm$ standard deviation; $P < 0.01$ by Student's two-tail t -test). Their higher rate of publication compared with co-authors who were not honorary appeared to depend entirely on their honorary papers: the difference disappeared when an author's honorary papers were not counted (0.67 ± 0.82 versus 0.71 ± 1.42). The 13 papers with at least one honorary author had more errors and discrepancies than the 5 without an honorary author (15.1 ± 10.2 versus 5.0 ± 6.6 ; $P = 0.067$).

Table 3 No. of dogs for which results were shared

	Ref. 53	Ref. 90	Ref. 91	Ref. 92	Ref. 94
Ref. 53	33	8	12	0	9
Ref. 90	8	22	9	0	11
Ref. 91	12	9	76	6	7
Ref. 92	0	0	6	48	0
Ref. 94	9	11	7	0	44

The boldface figures (on diagonal) give the total numbers of dogs used in each study. The remaining entries give the numbers of dogs for which experimental data were shared between pairs of studies. Although not shown in this table, data for the same 2 dogs were used in 4 studies^{53,90,91,94}. Reference 134 was the source of this table, which in turn was based on information provided by one of the senior co-authors at Harvard.

Incomplete statements

In many ways the most striking of our findings was the discovery that certain research papers (7 of 18) embody statements and data in a manner that impedes the reader in forming an accurate reconstruction of the way in which experiments were actually carried out. The circumstances were these.

Five papers^{53,90-92,94} reported studies done with dogs on the effects of experimental coronary artery occlusion on the heart and on recovery from the effects of occlusion with or without drugs. In these papers the assignment of dogs to experimental and control groups is described as formally randomized, as with the statement⁹⁰, "Twenty-five minutes after occlusion of the LAD, the remaining dogs were randomized to treated or control group on the basis of a random odd or even number generated by a computer program."

During the subsequent investigation by the NHLBI panel, however, the co-authors acknowledged that some control data had been reused¹³⁴ — in other words, that the control dogs actually included some historical controls, a use which had not been mentioned in the published papers. (Historical controls are animals that have been used as controls in a previous study¹³⁰. The same data recorded in the earlier study are then used again as control data in the later study. The number of dogs for which data were reused is shown in Table 3.)

One co-author of three of the five papers subsequently stated¹⁴⁶: "In order to economize animal usage, as well as professional and technical time, control animals which had previously been randomized were sometimes used to provide data for more than a single experiment . . . we saw no reason not to allow him [Darsee] to use his randomized controls in more than a single experiment." Another scientist, the co-author of all five papers, acknowledged the reuse of data, but stated¹⁴⁷ that Darsee had misled him and the other co-author about the number of dogs for which data were reused and about the number of experiments in which the data for any one dog were reused.

Reuse of data was approved by one of these co-authors in at least one paper not

involving Darsee (mentioned in ref. 134), and both co-authors have subsequently defended the practice. One co-author, commenting on a study not involving Darsee that was carried out in his laboratory, acknowledged¹⁴⁷ that control data were reused and that the reuse was not described on publication: "Although the use [of some control data from another study] was not specified in the text [of a study not involving Darsee], the use did not bias the results of the study in any way". The other co-author defended the sharing of control results said to be randomized by stating¹⁴⁶ that "the practice of not specifying in the Methods Section that a control dog had been used in more than one study, while certainly not ideal, is common," and also that "the practice of utilizing controls in more than a single experiment is not uncommon and, *although it should be*, it is not always spelled out in published papers" (emphasis added).

In a series of experiments in which the objective is to study the effects of various interventions on animals subjected to a certain procedure, in this case arterial occlusion, the ideal is that described in this group of five papers: all animals would be subjected to arterial occlusion, and afterwards assigned randomly either to the control group or to one or other of the experimental groups.

Although the use of historical controls may have several practical advantages, not least those of economizing in the use of animals, this practice is clearly less satisfactory than the ideal; for example, the experimenter's foreknowledge could systematically affect the outcome of experiments. The fact that Darsee may not have selected historical controls at random, as required by the laboratory's policy described to the NHLBI panel, is irrelevant to our criticism, which is that these papers leave the reader with the impression that control animals were selected by a randomization procedure that most readers, we judge, would consider superior to that actually used.

We emphasize that we do not criticize the use of historical controls, which may or may not have been appropriate, but merely the misleading published description. Thus the use of historical controls is reconcilable with the statement in the

paper quoted above that "after occlusion . . . the remaining dogs" were randomly assigned to different groups only if "remaining dogs" is understood to refer to a number of historical controls.

It would appear from the subsequent statements of the two co-authors that they knew of and authorized the use of historical controls in the laboratory, as well as the failure to describe that use on publication. The described randomization procedure in any study was supposed to apply to all dogs in that study, but in fact did not apply to the historical controls when their data were reused. We conclude that the papers contained an incomplete and misleading description of experimental procedures.

Furthermore, the experiments reported in the five papers included procedures, such as arterial catheterization and administration of certain drugs or dyes, that differ from one study to another (Table 4). As a result, most of the historical controls had been subjected to procedures in an earlier study that were different from those described for the later studies. Thus the published descriptions of these later procedures, which are described explicitly and in detail as applying to all dogs, do not apply in some respects to the historical controls.

One author, writing to the NIH about the sharing of results, stated¹⁴⁶ that "the experimental conditions were similar" for the studies in question. The other referred¹⁴⁷ to studies (not involving Darsee) in which "experimental conditions are the same"; he has also indicated in correspondence with us that the procedures described in the five papers were similar enough in some cases so that the use of historical controls was justified. Differences in experimental conditions are summarized in Table 4.

Some might consider that in most cases shown in Table 4, the data from dogs in one study would be unsuitable as reused data for another study, but the facts enabling readers to form a judgment on this point were not given in any of the five papers. We believe that the descriptions of experimental procedures did not apply to the reused data obtained in previous studies and that these descriptions were therefore misleading.

A final point may be related to reuse of data: one of the five papers states¹⁴⁸ that "all dogs [were subjected to an experimental procedure, which is described]. Ventricular fibrillation or standstill was then induced by a barbiturate overdose, and the hearts were excised rapidly and placed on ice." Yet the co-authors have said (cited in ref. 134) that some of the same dogs were subsequently used for the 4-hour and 6-hour occlusions in other experiments. As the NHLBI panel report politely observes¹⁴⁹, "This would appear to be impossible". We have assumed that the two co-authors in our sample were unaware of this anomaly at the time of publication.

The misleading statements noted above appeared in five papers co-authored by two members of our sample. In addition to these five papers, there were two others (refs 95 and 96; see next section) containing statements that were imprecise or misleading. In total, there were 18 instances of statements made by 6 co-authors whose effect would have been, in our opinion, to mislead the reader.

Unacknowledged republication

We were surprised to discover several instances in which the same data were republished without adequate reference to that circumstance. The most striking example was a paper⁹⁵ containing data later

published almost in their entirety as a major part of a subsequent paper⁹⁶ which appeared in a different journal. Although the second paper referred to the first by means of an index number in the text, there is no explicit statement in the second paper that the new data are a continuation of an experiment already published.

The subject matter of both papers concerns the recovery of heart muscle injured by temporary surgical occlusion of a coronary artery in dogs, described by the authors as a model with "important clinical implications for patients" with certain kinds of heart disease. The first paper states that "Seven dogs were also subjected to 15 min of LAD occlusion followed by 3 days of reperfusion", which is not false but which does not reveal that the same 7 dogs were actually reperfused for 14 days. The references in the second paper to the first would not suggest to those who were not familiar with the first that the research about to be described was a continuation of an experiment already published. In the second paper's opening paragraph, which is reproduced overleaf, the first paper is cited twice, but neither citation of the first paper provides a hint that some of the same data are being republished and that some of the same animals were used.

On neither paper was Darsee the first or last author, and it seems that the co-authors could have known of the statements that were imprecisely misleading. For the journal to which ref. 96 was submitted (*Proceedings of the National Academy of Sciences*), a contributing author customarily by-passes the refereeing system, which might have discovered the unacknowledged republication of data; instead the contributing author assumes "responsibility for the propriety and scientific standards"¹⁴⁸ of the manuscript he submits.

Table 4 Differences between experimental procedures in five papers

	Ref. 53	Ref. 90	Ref. 92	Ref. 94	Ref. 91
(1) Coronary artery occlusion (time)	2–6 h	6 h	1 or 3 h	4–6 h	1 min–2 h
(2) Randomization before/after coronary artery occlusion	After	After	Before	After	Before
(3) Time from removal of coronary artery clamp to end	~1 min–4 h	Not removed	5 or 72 h	Not removed or 2 h	6 h
(4) Antiarrhythmic agent:	Lido	Lido	Lido	Proc'de	None
(i) Initial bolus inj., mg per kg	1.5	1.5	1.5	3	None
(ii) Further intermittent inj.	Yes	No(?)	No(?)	No(?)	No
(iii) Continuous infusion	No	No	No	Yes	No
(5) Dye administered initially	Flu	MeB, GV	Flu, GV	Flu	Flu, GV
(6) Dye injected terminally	TS, MeB	MeB	MeB	MonB	MonB
(7) Drug or NaCl soln, ml per kg	0(?)	1	1	2	0(?)
(8) Catheter in aorta/left carotid	Carotid	Carotid	Carotid	Aorta	Carotid
(9) A set of multiple procedures, (a)–(e), listed below	No	No	Yes	No	Yes

There are substantial differences in experimental procedures for at least nine of the ten pairs of papers. The procedures in item 9 include: (a) insertion of micromanometer through stab wound in ventricular apex, (b) application of coronary cuff, (c) repeated brief reocclusion of coronary artery, (d) implantation of crystal transducers in myocardium, and (e) delay in examination of heart after death (for flowmeter calibration). Abbreviations: Lido, lidocaine; Proc'de, procainamide; Flu, fluorescein (injection); MeB, methylene blue (injection); GV, gentian violet (topical); TS, thioflavine S; MonB, monastral blue. Information in the table comes from the papers themselves^{53,90–92,94}; see also ref. 134, appendix 5.

Opening paragraph of ref. 96:

PREVIOUS studies from this laboratory have shown that brief (15 min) coronary occlusions followed by reperfusion do not cause necrosis but do result in prolonged abnormalities in myocardial biochemistry, function, and ultrastructure (1–3). At 72 h after release of a 15-min coronary occlusion, the ATP concentration in reperfused previously ischemic myocardium was significantly reduced (by 22% of normal), the percentage of systolic shortening of reperfused left ventricular segments was reduced (by 42% of normal), and ultrastructural abnormalities were present, despite the absence of necrosis (1, 3). However, it was not known whether these changes ever were reversed to normal after longer periods of reperfusion and, if so, what the time course of the recovery is. Recently there has been renewed interest in the treatment of acute myocardial infarction by means of coronary reperfusion, either by injecting fibrinolytic agents directly into the obstructed coronary artery or by coronary revascularization (4, 5).

This is the opening paragraph, printed here verbatim, of a research paper⁹⁶, discussed in the text. The two occurrences of reference 3 in the paragraph above are the only references in this paper⁹⁶ to the authors' own previous study⁹⁵ (see text).

A common example of republication involves abstracts of research submitted to scientific societies. Of the 88 abstracts, only 47 were published only once; all the rest involve duplicate or triplicate publication. This practice is apparently followed in many fields of biomedical science, where an abstract may be prepared for one meeting of a society and then submitted, perhaps in an amended form, to a second meeting. We nevertheless suspect that the extent of this practice may not be widely appreciated. We discuss four cases below but have not included abstract data in our tables.

We note that there were four pairs of Darsee abstracts in which the texts were nearly identical, but whose titles had apparently been changed. For example, an abstract⁷⁷ entitled "Can tetrazolium stains reliably identify myocardial infarction prior to definite histologic necrosis?" was also presented⁷³ under the title "Early pathologic detection of acute myocardial infarction". Sometimes the alteration consists of the substitution of synonyms^{80,72}: "Persistent myocardial abnormalities following brief periods of temporary coronary occlusion not associated with necrosis" also appears as "Prolonged metabolic, functional, and ultrastructural abnormalities following transient myocardial ischemia without infarction".

In each pair the authors are the same, and the text and numerical values are identical through most of the abstract. The publication dates were generally within about a month of one another, as were the scheduled presentation dates.

Although it could be argued that it may be necessary to tailor the titles of certain abstracts to make them acceptable for oral presentation before the different societies, we consider that the change in title has the effect of making the pair of abstracts, if listed only by title, appear different to the reader, whereas in fact the contents are almost the same.

There were 5 instances of unacknowledged republication of data. They appeared in publications, two from Emory and seven from Harvard, having a total of 8 co-authors (Table 2).

Lapses from standards

We were surprised to find that there is apparently no generally accepted written code of conduct for scientific research¹⁴⁹⁻¹⁵⁹. We think that there is nonetheless an unwritten code of which scientists are aware and to which they profess adherence. We have based the analysis that follows on standards that we believe are generally accepted—accepted in the sense that they are generally understood, not necessarily in the sense that they are always complied with.

In Table 5 we attempt to place in ten categories the departures from generally accepted standards that we observed in our sample; other categories would probably be more appropriate for other studies. We found it useful to divide our categories into two groups, A and B, with those of group A explicable simply by carelessness or excessive haste, while those of group B appeared to be more serious in some sense. We do not mean to suggest that certain departures from accepted practice which appear in group B may not be due solely to carelessness and haste; undoubtedly this is so in some cases. Furthermore, we recognize that the distinction between type A and type B is necessarily subjective and that other observers might classify the lapses we have found differently.

We emphasize that the instances on

which Fig. 4 is based were supported by evidence we considered compelling, but not necessarily conclusive in each case. We acknowledge that our categories must be subjective.

Frequency of lapses

Of the 47 co-authors in our sample, 31 (66%) were authors of publications containing lapses of type A—a result that is the same whether or not honorary authorships are included. Furthermore, 13 (28%) were involved in type B lapses. Only 12 of the 47 were found not to have been involved in lapses of either type. Most of the 12 were authors of abstracts only; just 2 of the 12 were co-authors of a research paper.

Some scientists apparently do not regard abstracts as publications, even though they are archived and cited and are sometimes used to establish priority¹⁶⁰. If the 88 abstracts and 3 chapters are excluded from consideration, the frequency of type A and B lapses is about the same or higher. Of the 22 scientists who were co-authors of the 18 research papers, 21 (95%) were involved with papers containing lapses of type A, and 6 (27%) were involved in lapses of type B.

Discussion

It is obvious that the lapses from generally accepted standards of research described here have at least the potential for interfering with the accuracy of the scientific literature or for harming the scientific enterprise in some other way. We also think that the instances described above represent departures from the generally accepted standards of scientific research. We recognize that the decision as to what constitutes a departure from generally accepted standards is necessarily subjective. Presumably some scientists have views different from those expressed here. We regard it as a matter of the greatest importance that these differences be discussed publicly.

Table 5 Lapses from generally accepted standards

Type A

- (1) Presence of errors (usually in numerical values).
- (2) Inconsistency with a research group's previously published data.
- (3) Failure to retain names or identifying numbers for human subjects included in a published paper.
- (4) Honorary authorship (see text).

Type B

- (5) Statements that are misleading.
- (6) Misleading citations of a previous paper containing data being republished, so that the true relationship between the two papers is obscured.
- (7) Publication of very similar abstracts under very different titles.
- (8) Failure to utilise unique knowledge that makes possible the recognition of certain serious errors in published work—errors not discoverable by anyone else.
- (9) Failure to acknowledge the source of a substantial amount of research data received from someone else.
- (10) Failure to take appropriate action after receipt of a complaint, later shown to be well founded, that a colleague may be involved in questionable data collection.

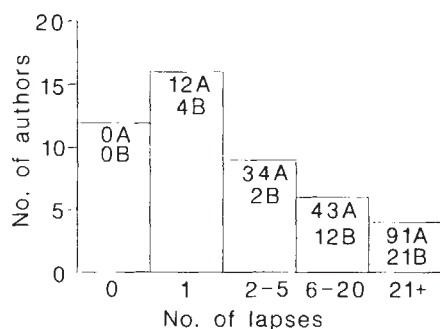


Fig. 4 Authors and their lapses. The histogram gives the number of authors in our sample having lapses within the specified range. The numbers give the total of type A and type B lapses for all authors in that range.

The abundance of errors that we observed in our sample obviously interferes with the accuracy of the scientific literature. Every scientist has, at a minimum, an obligation to ensure that what is published under his name is accurate.

There is a notable absence of explicit published standards for the retention of data. We believe that it would be generally accepted that data should remain accessible for as long as questions and criticisms are likely to be raised about what has been published. For most of the papers in our sample, data were not available. The data we give in Table 4 use only those cases where data involving human subjects had not been retained. We suggest that a reasonable minimum requirement is that each author of a paper involving human subjects should retain at least a list of the subjects' names or the hospital's identification numbers of patients. This information would permit access to hospital records and thus in many cases allow reconstruction of data in the paper.

The reader may ask: What harm is done by honorary authorship? Indeed some of our colleagues have argued that the custom of routinely placing the name of a senior scientist, usually the head of the laboratory, on a paper — regardless of his contribution — is widely followed and does no harm. We disagree, as have others^{159,161-166}: honorary authorships falsify the assignment of responsibility for published research and increase the likelihood that inaccurate data will be published. The honorary author is in a poor position to judge the validity of the work, yet he often lends a prestige that may lull other co-authors, the reviewers or the readers into uncritical and inappropriate acceptance.

Similarly, the harm done by unacknowledged republication of data may not be immediately obvious. But it disguises the true relationship between published papers and yields no benefits other than for the authors.

In some ways the most disturbing of our findings is the discovery that certain of the publications (7 out of 18) embody state-

ments and data in a manner that impedes the reader in forming an accurate reconstruction of the way in which experiments were carried out.

Representativeness

The evidence presented here indicates a surprisingly high frequency of lapses from accepted standards. One possible explanation is the association of the co-authors of the publications in our sample with Darsee: he may have induced them to behave in ways not typical of them. A second possibility is that the co-authors are not representative of their field. A third possibility is that lapses from accepted standards may be unusually common in the specialty in which these scientists worked, in clinical research generally or in research carried out at medical schools or by physicians. (The co-authors of the publications in our sample were mainly physicians, and the research was carried out exclusively in medical schools.) A fourth and disturbing possibility is that lapses from accepted standards may be more common among biomedical scientists than is currently appreciated.

Other studies will be needed to determine whether our results are representative of other groups of scientists.

Causes

Extraordinary emphasis is commonly placed on the sheer quantity of publications^{158,165-179}. The director of one of the world's leading research institutions (but not an institution with which Darsee was ever affiliated) sent a signed, official memorandum to junior colleagues proposing incentives for publication: "There is no demand that these be literary masterpieces in first line journals; journeyman works for publication in second, third or fourth line archival publications will be quite satisfactory. Upon proper completion and submission of [two] manuscripts, [a technician]'s appointment will be extended to April 1, 1984. During that time it is expected that an additional manuscript on [subject] will be completed and submitted. If so, the period of employment will be extended an additional three months and again an additional manuscript on [subject] is an anticipated result of the extended employment." On another occasion a director of research said in all seriousness, "No research that results in a single paper is worthwhile."

Those promoting Darsee or recommending his promotions were well aware that his publications were unusually numerous¹⁶⁶. But apparently none of these supporters, who were also his co-authors, made the effort that would have shown that many of the papers had serious flaws. How often does this occur?

Other causes of lapses from accepted standards may be the extreme competition for research grants, large research

teams in which the leader is not directly involved in the research and the absence of an accepted code of conduct for scientific research.

In this paper, we have deliberately avoided a discussion of the cases of acknowledged fraud which have come to light in recent years, concentrating instead on defects of the scientific literature generally regarded in the scientific community as regrettable but less serious. Outright fraud, which might conveniently be called a lapse of type C in our nomenclature, is presumably rare. May not the scientific community be over-sanguine in attaching less importance to lapses such as we have found, which are generally unremarked but which debase the scientific literature?

Testing the system

The results of our study naturally raise the question: what fraction of papers in the biomedical sciences are not supported by primary data at the time of publication? The question cannot be answered with the papers in our sample, since most are known to be partial or complete forgeries and thus are atypical.

We suggest a study to answer the question: close examination of a random sample of published papers. Probability of selection could be uniform or could be weighted by a variable such as number of authors or cost of research. Papers would be examined by either of two methods: careful examination of the paper itself for errors and discrepancies by appropriate experts following an established procedure ("external audit") or examination of the primary data on which the paper is based, which would of course require access to laboratory records ("internal audit"). The advantages of the external audit are its non-invasive character and presumably lower cost; the advantage of the internal audit is its completeness.

What would this cost? This depends on the precision required and on the actual fraction of defective papers. We assume for the sake of illustration that an external audit could be done for about 3% of the cost of the research on which a paper is based, and an internal audit for about 10%. If the actual fraction of papers that could not be substantiated were as high as 15% and if the answer were required with a standard error of 7%, about 25 papers would have to be examined, which in the case of internal audits would altogether cost only about 2.5 times as much as one average paper. If, on the other hand, auditing 100 papers failed to disclose a single paper that could not be substantiated, this would show with $P = 0.05$ that the actual fraction of papers unsupported by primary data at the time of publication is less than 3%.

In addition to the financial costs of examining the practices of scientists, there are other costs which may be more serious

in the long run. Examination of scientific practices could cause unwarranted harm to individual scientists. Systematic examination of scientific practices might even weaken the fabric of trust that is essential to the functioning of science. Of all human endeavours, science is one of the most successful — prodigious in benefits, low in cost. But science, vulnerable to abuse from within by its practitioners, is perhaps even more vulnerable to harm by regulation, and at some point the cost of further regulation will outweigh the benefits.

Scientists have to an unusual degree been entrusted with the regulation of their own professional activities. Self-regulation is a privilege that must be exercised vigorously and wisely, or it may be lost. □

Walter Stewart and Ned Feder are in the National Institutes of Health, Bethesda, Maryland 20892, USA.

1. *Notre Dame Sci. Q.* 7, no. 2, 14-17 (1969).
2. *Notre Dame Sci. Q.* 8, no. 1, 9-11 (1969).
3. *Circulation* 56, III-218 (1977).
4. *South. med. J.* 70, 1249 (1977).
5. *Archs intern. Med.* 137, 1639 (1977).
6. *Ann. intern. Med.* 89, 867-870 (1978).
7. *Clin. Res.* 26, 3A (1978).
8. *Clin. Res.* 26, 226A (1978).
9. *Clin. Res.* 26, 226A (1978).
10. *Chest* 74, 334-335 (1978).
11. *Chest* 74, 334 (1978).
12. *Circulation* 58, II-237 (1978).
13. *Circulation* 58, II-235 (1978).
14. *Circulation* 58, II-229 (1978).
15. *Circulation* 58, II-174 (1978).
16. *Circulation* 58, II-42 (1978).
17. *Circulation* 58, II-23 (1978).
18. *Clin. Res.* 26, 773A (1978).
19. *Clin. Res.* 26, 762A (1978).
20. *Clin. Res.* 26, 759A (1978).
21. *Clin. Res.* 26, 745A (1978).
22. *New Engl. J. Med.* 298, 221-222 (1978).
23. *Am. J. Med.* 64, 1089 (1978).
24. *Circulation* 59, 492-497 (1979).
25. *Circulation* 59, 619-622 (1979).
26. *New Engl. J. Med.* 300, 877-882 (1979).
27. *Am. J. Cardiol.* 43, 367 (1979).
28. *Clin. Res.* 27, 274A (1979).
29. *Clin. Res.* 27, 265A (1979).
30. *Clin. Res.* 27, 264A (1979).
31. *Clin. Res.* 27, 230A (1979).
32. *Clin. Res.* 27, 161A (1979).
33. Abstract of report presented in annual Lloyd Hyde Medical House Staff and Fellows Research Program, Emory University School of Medicine (1979).
34. *Clin. Res.* 27, 573A (1979).
35. *Clin. Res.* 27, 561A (1979).
36. *Clin. Res.* 27, 561A (1979).
37. *Chest* 76, 371 (1979).
38. *Chest* 76, 350 (1979).
39. *PACE* 2, A43 (1979).
40. *Circulation* 60, II-157 (1979).
41. *Circulation* 60, II-149 (1979).
42. *Circulation* 60, II-124 (1979).
43. *Circulation* 60, II-90 (1979).
44. *Circulation* 60, II-41 (1979).
45. *Circulation* 60, II-38 (1979).
46. *Clin. Res.* 27, 726A (1979).
47. *New Engl. J. Med.* 301, 443 (1979).
48. *South. med. J.* 72, 174-180 (1979).
49. *Circulation* 60, II-252 (1979).
50. *Ann. intern. Med.* 92, 735-741 (1980).
51. *Am. J. Cardiol.* 46, 607-612 (1980).
52. *Am. J. Cardiol.* 46, 613-618 (1980).
53. *Am. J. Cardiol.* 46, 800-806 (1980).
54. *Am. J. Cardiol.* 45, 492 (1980).
55. *Am. J. Cardiol.* 45, 476 (1980).
56. *Am. J. Cardiol.* 45, 395 (1980).
57. *Clin. Res.* 28, 257A (1980).
58. *Clin. Res.* 28, 163A (1980).
59. *Clin. Res.* 28, 163A (1980).
60. *Clin. Res.* 28, 653A (1980).
61. *Chest* 78, 513 (1980).
62. *Chest* 78, 513 (1980).
63. *Circulation* 62, III-318 (1980).
64. *Circulation* 62, III-317 (1980).
65. *Circulation* 62, III-177 (1980).
66. *Am. J. Cardiol.* 45, 394 (1980).
67. *Fedn Proc.* 39, 1112 (1980).
68. *Fedn Proc.* 39, 398 (1980).
69. *Clin. Res.* 28, 538A (1980).
70. *Clin. Res.* 28, 163A (1980).
71. *Clin. Res.* 28, 174A (1980).
72. *Clin. Res.* 28, 615A (1980).
73. *Clin. Res.* 28, 614A (1980).
74. *Clin. Res.* 28, 610A (1980).
75. *Clin. Res.* 28, 609A (1980).
76. *Circulation* 62, III-314 (1980).
77. *Circulation* 62, III-246 (1980).
78. *Circulation* 62, III-239 (1980).
79. *Circulation* 62, III-196 (1980).
80. *Circulation* 62, III-80 (1980).
81. *Circulation* 62, III-76 (1980).
82. In *Clinical Methods: The History, Physical and Laboratory Examinations* (eds Walker, H. K., Hall, W. D. & Hurst, J. W.) 1122-1161 (Butterworth, Boston, 1980).
83. In *The Heart, Update III* (ed. Hurst, J. W.) 1-8 (McGraw-Hill, New York, 1980).
84. In *Heart Disease: A Textbook of Cardiovascular Medicine*, Vol. 2 (ed. Braunwald, E.) 1517-1582 (Saunders, Philadelphia, 1980).
85. *Primary Cardiol.* 6, 101-109 (1980).
86. *Primary Cardiol.* 6, 112-118, 122 (1980).
87. *New Engl. J. Med.* 304, 129-135 (1981).
88. *Ann. Surg.* 194, 189-192 (1981).
89. *Circulation Res.* 49, 1017-1028 (1981).
90. *Circulation* 63, 29-35 (1981).
91. *Am. J. Physiol.* 240, H399-H407 (1981).
92. *J. clin. Invest.* 68, 225-239 (1981).
93. *Archs Path. Lab. Med.* 105, 403-406 (1981).
94. *Am. J. Cardiol.* 48, 702-710 (1981).
95. *Am. J. Physiol.* 241, H591-H599 (1981).
96. *Proc. natn. Acad. Sci. U.S.A.* 78, 7152-7156 (1981).
97. *Clin. Res.* 29, 444A (1981).
98. *Clin. Res.* 29, 428A (1981).
99. *Chest* 80, 388 (1981).
100. *Circulation* 64, IV-280 (1981).
101. *Circulation* 64, IV-210 (1981).
102. *Circulation* 64, IV-69 (1981).
103. *Circulation* 64, IV-68 (1981).
104. *Fedn Proc.* 40, 485 (1981).
105. *Clin. Res.* 29, 562A (1981).
106. *Clin. Res.* 29, 270A (1981).
107. *Clin. Res.* 29, 270A (1981).
108. *Clin. Res.* 29, 184A (1981).
109. *Clin. Res.* 29, 184A (1981).
110. *Clin. Res.* 29, 184A (1981).
111. *Z. Kardiologie* 70, 317 (1981).
112. *J. nucl. Med.* 22, P53 (1981).
113. *Chest* 80, 385-386 (1981).
114. *Circulation* 64, IV-280 (1981).
115. *Circulation* 64, IV-236 (1981).
116. *Circulation* 64, IV-222 (1981).
117. *Circulation* 64, IV-178 (1981).
118. *Circulation* 64, IV-122 (1981).
119. *Circulation* 64, IV-99 (1981).
120. In *The Heart, Update IV* (ed. Hurst, J. W.) 219-286 (McGraw-Hill, New York, 1981).
121. *Am. Heart J.* 102, 807-808 (1981).
122. *Am. Heart J.* 101, 124-126 (1981).
123. *J. cardiovasc. Med.* 6, 705-715 (1981).
124. *Acta med. scand., Suppl.* 651, 123-132 (1981).
125. *J. cardiovasc. Med.* 6, 1173-1177, 1181 (1981).
126. *Am. J. Cardiol.* 49, 1023 (1982).
127. *New Engl. J. Med.* 308, 1419 (1983) (retraction).
128. *Ann. intern. Med.* 99, 275-276 (1983) (retraction).
129. *Circulation Res.* 53, 837 (1983) (retraction).
130. Last, J. M. (ed.) *A Dictionary of Epidemiology*, 23, 94 (Oxford University Press, New York, 1983).
131. Ross, R. S. et al. Report of an ad hoc advisory committee to the dean of the Harvard Medical School on dishonesty in scientific research, *Harvard University Gazette* 77, 1, 11-12 (29 January 1982).
132. Hall, E. C. et al., Report of ad hoc committee to evaluate research of Dr. John R. Darsee at Emory University; [the "Moran report"]; 1-32 (Emory University, 28 March 1983); cf. *Minerva* 23, 276-304 (1985).
133. Moran, N. C., Addendum to report on research of Dr. John R. Darsee at Emory University School of Medicine, 1-3 (Emory University, 5 May 1983).
134. Morgan, H. E., Klocke, F. J., Shepherd, J. T., & Wildenthal, C. K., Report of the National Heart, Lung, and Blood Institute special panel to review alleged misconduct at Brigham and Women's Hospital/Harvard Medical School, Boston, Massachusetts, 1-36 and 7 Appendices (29 June 1982).
135. Knox, R. J. *Am. med. Ass.* 249, 1797-1807 (1983).
136. Knox, R. A. *J. Am. med. Ass.* 249, 2867-2876 (1983).
137. Knox, R. A. *Boston Globe*, 41, 46 (23 May 1983).
138. Broad, W. J. *Science* 215, 478-482 (1982).
139. Broad, W. J. *Science* 215, 874-876 (1982).
140. Culliton, B. J. *Science* 220, 31-35 (1983).
141. Culliton, B. J. *Science* 220, 936 (1983).
142. Heymsfield, S. B. & Glenn, J. F. *New Engl. J. Med.* 308, 1400 (1983).
143. [Braunwald, E.] *Proc. natn. Acad. Sci. U.S.A.* 79, 6390 (1982).
144. Milnor, W. R. *Hemodynamics*, 17 (Williams and Wilkins, Baltimore, 1982).
145. *Science Citation Index, Source Index* (Institute for Scientific Information, Philadelphia, 1979-86).
146. Braunwald, E., "Comments concerning NHLBI panel's report and range of possible NIH actions", Memorandum to NIH staff members, 1-36 and 6 Appendices (21 September 1982).
147. Kloner, R. A., Memorandum to W. F. Raub, NIH, 1-21 (14 September 1982).
148. "Information To Contributors", *Proc. natn. Acad. Sci. U.S.A.* 78, ix (January 1981).
149. Krevans, J. R. et al. *J. med. Educ.* 57, 895-902 (1982).
150. Pigman, W. & Carmichael, E. B. *Science* 111, 643-647 (1950).
151. Clapp, J. *Professional Ethics and Insignia*, 1-851 (Scarecrow Press, Metuchen, New Jersey, 1974).
152. "Yale Policy Statement on Collaborative Research", *Yale Weekly Bull. Calendar* 11, 9 (13-20 September 1982).
153. Hastings Center Staff (eds) *The Hastings Center's Bibliography of Ethics, Biomedicine, and Professional Responsibility*, 10-11 (University Publications of America, Frederick, Maryland, 1984).
154. Danforth, W. H. et al. *Report of the Association of American Universities Committee on the Integrity of Research*, 1-6 (Association of American Universities, Washington, DC, [April 1983]).
155. *The Chemist's Creed*, American Chemical Society (ACS, Washington, DC, 14 September 1965).
156. *Honor in Science*, 1-38 (Sigma Xi, New Haven, Connecticut, 1984).
157. Chalk, R., Frankel, M. S. & Chafer, S. B. *Professional Ethics in the Scientific and Engineering Societies*, 1-224 (American Association for the Advancement of Science, Washington, DC, 1980).
158. Huth, E. J. *Ann. intern. Med.* 104, 266-267 (1986).
159. Huth, E. J. *Ann. intern. Med.* 104, 269-274 (1986).
160. Stossel, T. P. *New Engl. J. Med.* 313, 123-126 (1985).
161. Cannon, W. B. *The Way of an Investigator: A Scientist's Experience in Medical Research*, 93 (Norton, New York, 1945).
162. [Huth, E. J.] *Ann. intern. Med.* 97, 613-614 (1982).
163. Reiman, A. S. *New Engl. J. Med.* 308, 1415-1417 (1983).
164. Croll, R. P. *Perspect. Biol. Med.* 27, 401-407 (1984).
165. Majerus, P. W. *J. clin. Invest.* 70, 213-217 (1982).
166. Huth, E. J. *Ann. intern. Med.* 104, 257-259 (1986).
167. Broad, W. & Wade, N. *Betrayers of the Truth* (Simon & Schuster, New York, 1982).
168. [Huth, E. J.] *Ann. intern. Med.* 99, 266-267 (1983).
169. Burman, K. D. *Ann. intern. Med.* 97, 602-605 (1982).
170. Gjerde, C. L. & Colombo, S. E. *J. med. Educ.* 57, 157-162 (1982).
171. Angell, M. *Science* 219, 1417-1418 (1983).
172. Mahoney, M. J. *Scientist as Subject: the Psychological Imperative* 79-82 (Ballinger, Cambridge, Massachusetts, 1976).
173. Jackson, C. I. & Prados, J. W. *Am. Scientist* 71, 462-464 (1983).
174. Gratzner, W. *Nature* 302, 774-775 (1983).
175. Stetten, D., Jr. *Science* 226, 1374, 1376 (1984).
176. Petersdorf, R. G. *Ann. intern. Med.* 104, 252-254 (1986).
177. Woolf, P. K. *Ann. intern. Med.* 104, 254-256 (1986).
178. Bailar, J. C., III. *Ann. intern. Med.* 104, 259-260 (1986).
179. Angell, M. *Ann. intern. Med.* 104, 261-262 (1986).

● Some editorial changes have been made in this manuscript without the consent of the authors. A reply from Braunwald follows on page 215.

On analysing scientific fraud

from Dr Eugene Braunwald, Harvard Medical School

ANYONE who has been a victim of scientific fraud or who has had close contact with such an episode must be sobered by the event and must ask how the fraud could have been prevented or detected earlier. The costs are incalculable, not only for the persons and institutions involved, but for science as a whole. An objective analysis of research practices which offered some insight into the nature of scientific fraud and the manner in which it could be prevented or detected would be an important contribution to the public discussion of this critical issue.

I agree that many of the issues Stewart and Feder raise are important and deserve attention and thoughtful analysis. When specific incidents of scientific misconduct are discussed, however, it is essential that the facts be dealt with accurately and objectively. Unfortunately, Stewart and Feder do not provide such a discussion of the events surrounding the Darsee case insofar as they occurred at the Cardiac Research Laboratory at Harvard and the Brigham and Women's Hospital.

As the paper has been widely circulated in many different versions and its contents have been released to the lay press before publication, I am pleased to have the opportunity to reply to the paper with respect to the work done by the 23 coauthors at Harvard.

John Darsee committed serious fraud in at least three (not two) universities over a period of at least 12 years. The premise implicit in Stewart and Feder's paper is that it is possible to separate clearly *this* fraud committed by Darsee from the professional behaviour of his coauthors. In fact, the paper repeatedly and unfairly connects Darsee's fraud at Harvard to his coauthors there through a process of guilt by association. As a result, many of Stewart and Feder's allegations are misattributions of Darsee's wrong doing to his coauthors. Other allegations involving the papers from Harvard — all of which were retracted in 1982 — are insubstantial or simply wrong and have nothing to do with scientific misconduct.

Allegations of errors

Stewart and Feder assert that there were errors in the Harvard papers although the specifics and location of those purported errors in the papers have not been provided despite numerous requests. This assertion is highly dubious as the few alleged errors which Stewart and Feder actually identify prove, upon careful scrutiny, not to be errors at all or to be wholly trivial errors which most scientists would agree do not rise to the level of a "lapse" in research standards.

Stewart and Feder discuss one Harvard paper, Paper 96, in which they purport to have identified five errors. The first involves a supposed semantic inconsistency

concerning when systolic shortening of the previously ischaemic heart returned to normal. The legend to a figure in Paper 96 stated that recovery of systolic shortening did not occur until *after* 7 days, but Stewart and Feder claim that the figure and a table reflected recovery *at* 7 days. Paper 96 makes clear that the data presented in the figure and the table were obtained at the end of 7 days and the phrase "after 7 days" should not mislead anyone. If this represents an error in phrasing at all, I agree with Stewart and Feder that it is of "very minor significance".

Equally insignificant are two claimed discrepancies between values reported in the summary and text of Paper 96, which Stewart and Feder concede are "small". Whether these small differences resulted from the different methods by which these figures could be calculated (on an experiment-by-experiment basis versus a group basis) or from typographical errors, they did not affect the findings of Paper 96 and would not mislead any reader. The general understanding of scientific fraud is hardly advanced by a discussion which hinges upon typographical errors and similar quibblings.

Stewart and Feder also point to a supposed discrepancy between the value of negative dp/dt for one dog in Paper 96 and for six dogs in Table 2 of Paper 95, which had been submitted for publication six months earlier. These alleged errors can be identified only with hindsight, for there is nothing inherently suspicious about the error bars or the dp/dt values. Although these may ultimately be inaccurate as a result of Darsee's fraud, it was not a questionable practice for the Harvard coauthors *at the time* to have accepted Darsee's calculations. Any other course would require coauthors of each paper to recalculate individually every value reported, denying researchers the considerable benefits of a reasonable division of labour. Indeed, in the type of experiment reported, although the absolute values of the cardiovascular variables may differ considerably among different animals, they tend to change in a similar manner over time, and it should not be surprising that the error bars are similar.

Fourthly, the purported discrepancy between the number of experiments in the two papers and the similarity of the results relates to the fact that the number of animals studied at various times differed. The results in those animals reperused for the shortest period were similar in the two papers. Smaller numbers were reperused for longer periods. Stewart and Feder's suggestion that those values and error bars could not be correct is unfounded. At most, there was ambiguity about which dogs provided biochemical data at different times. This is not a "lapse of standard".

Lastly, Stewart and Feder claim an inconsistency between an error bar in a figure and the numerical values presented in the text and summary. That and other figures in Paper 96 reflect numbers in the text within the technical limitations of artwork in scientific papers.

Retention of data

I certainly agree with Stewart and Feder about the need to retain research data, but they convey a totally inaccurate picture about this practice in the Harvard laboratory because they fail to mention that the NHLBI panel expressly found that only Darsee failed to retain primary data, and that the other Harvard researchers did retain original records. The NHLBI report stated, "With the exception of Dr Darsee, recent trainees [in the Harvard laboratory] have retained original records in customary and acceptable fashion". The laboratory policy on retention of data was clear; Darsee's failure to retain data was in direct contravention of the policy. Stewart and Feder would no doubt agree that it is proper practice in describing results to provide all of the relevant information. They did not have direct information about retention of data; the NHLBI Panel did. It is unfair to criticize the Harvard coauthors about this issue while withholding from the reader relevant information obtained by the panel.

Involvement in research

I fully agree with Stewart and Feder's general conclusion about the corrosive influence of "honorary coauthorship", but, as they admit, their assessment of this issue is based on subjective inferences drawn from "limited" evidence. Had they had access to all of the relevant facts, I believe they would have known that there were no honorary coauthors of Darsee's Harvard papers. I developed the overall research strategy — a continuation of a research programme begun in 1968, wrote the grant proposals which supported much of the work, participated actively in the design of the protocols, reviewed the results on a frequent, usually biweekly basis and participated actively in the interpretation of the data. This amply satisfies stringent requirements for coauthorship. Indeed, failure to designate a collaborator with this level of involvement as a coauthor would be misleading. (Had I not been included as a coauthor, Stewart and Feder would have been justified in criticizing me for not shouldering my share of the responsibility.) How is the cause of science served by making such unfair accusations?

Randomization

Stewart and Feder's discussion of the manner in which control dogs were selected is predicated on a misunderstanding

ing. The laboratory had a policy for randomizing control dogs, as follows: when a series of interventions were studied on myocardial infarct size in dogs subjected to coronary occlusion, a comparison was made between infarct size in treated and control dogs. The latter were selected in the following manner: with the first intervention, all control dogs were selected at random and were studied *de novo* and contemporaneously with the treated dogs. With subsequent interventions, when the research fellow was the same and the experimental conditions similar, some control dogs were studied *de novo*, that is, contemporaneously with the treated dogs, whereas the remainder were selected, again at random, from among the controls which had been obtained at random by the same research fellow a short time before.

This practice saved unnecessary sacrifice of animals, but allowed each intervention to be compared to a valid control group. As we advised the NHLBI panel in 1982, it would have been preferable to present this policy explicitly in the papers, but this would not have influenced the results or conclusions. Like all experimental procedures, this ultimately relied on the honesty and integrity of the individual researchers. What did influence the results is that Darsee violated this policy.

Deeply troubling is that Stewart and Feder quote selectively my statements concerning the laboratory's policy of using control dogs. By specifically excluding from the quotation references to key aspects of laboratory policy, they permit the reader to be misled that the Harvard coauthors knew and sanctioned Darsee's dishonest practices in this regard.

Republication

Stewart and Feder allege four instances of "unacknowledged republication" of data from Harvard, one involving two research papers. The first described successful experiments on 21 dogs which were studied 90 minutes and 72 hours following reperfusion after brief coronary occlusion; 14 of these 21 dogs were sacrificed 72 hours after reperfusion and heart tissue was analysed for histology, biochemistry and electron microscopy with the results reported in this paper. Physiological data obtained 72 hours after reperfusion in seven dogs were also reported in this paper; these seven dogs were not sacrificed. It was deemed necessary to apprise the scientific community of the findings, which had important clinical implications. Therefore, we decided to publish these results as the work continued and additional data became available.

The second paper included substantial new information, including the results of successful experiments on 20 additional dogs. That paper reported biochemical, electronmicroscopic and histochemical data on these 20 additional dogs, which were studied either 7 or 14 days following reperfusion. The second paper also reported for the first time: (1) physiological data obtained 7 and 14 days follow-

ing reperfusion in the 7 dogs for which data obtained 72 hours after reperfusion had been described in the first paper and which were not sacrificed and (2) histochemical data on dogs sacrificed at 72 hours. The second paper recapitulates the earlier findings, making express reference to the first paper twice in the opening paragraph. Would we have done this if we had wished to conceal republication?

Stewart and Feder suggest that there were lapses in standards in the submission of certain Harvard abstracts. However, these abstracts were prepared in strict conformity with the rules of the sponsoring societies. None was submitted for duplicate presentation at national meetings. To suggest that there is something improper in revising an abstract to enhance its quality and clarity is baffling. For example, Abstract 70 was submitted but not accepted for presentation to the American Federation for Clinical Research. After the abstract was revised to improve its clarity and focus, it was accepted as Abstract 78 for presentation seven months later at the scientific session of the American Heart Association. Stewart and Feder's contention that the Harvard coauthors somehow engaged in a questionable practice when they revised a previously rejected abstract before re-submitting that work flies in the face of commonly accepted practice.

Conclusions

Just as it is incumbent on scientists to be meticulous about the manner in which they carry out their work, so it is incumbent on analysts of the conduct of science to be meticulous about their activities. Just as it is damaging to science to make exaggerated claims not supported by data, so is it damaging to science to make exaggerated criticisms and allegations about the conduct of science which are not fully supported by facts. Just as scientists must report the results of all relevant experiments and not select just those which support their conclusions while deleting those which do not, so should analysts of science adhere to the same rules.

Despite my differences with Stewart and Feder's analysis regarding Darsee's work at Harvard, the larger problem which they address, albeit obliquely, namely that of fraud in science, is regrettably still with us and must not be swept under the rug. When we discovered and ultimately exposed all of Darsee's misconduct we were staggered at its extent. Nothing in my more than 25 years in research prepared me to deal with it. In retrospect, there are aspects of the matter that could have been handled better. Perhaps one positive outcome of the unfortunate Darsee affair is that it has helped to focus attention on fraud in science, in part prompting both research institutions and granting agencies to develop more formal rules and procedures for dealing with suspected fraud.

I have thought much about fraud in science during the past five and a half

years. Although I have not found any simple, new solutions, a number of issues are clearer to me now.

First, it must be understood by all that to be a scientist is a privilege and that society invests a special trust in all scientists. Whenever that trust is abused it diminishes all scientists. Second, because the costs of serious fraud are so high, we must be more vigilant and more sceptical, always alert to the possibility, however remote, that data we have been shown have been misrepresented. Third, in awarding promotions and grants, undue weight is sometimes given to the quantity as opposed to the quality of a scientist's contributions. This unfortunate practice may lower the quality of science, but I do not believe that it is the root cause of fraud. Fourth, it is usually not possible to explain the deeper motivation for the commission of serious scientific fraud within the framework of normal behaviour. I believe that many instances of scientific fraud represent a form of unconscious self-destructive behaviour which may have aggressive components directed against the perpetrator's supervisors, colleagues and institution; indeed, it mocks science as a whole. Fifth, once a single act of outright dishonesty is discovered, the burden should shift to proving that other data presented by the scientist in question were obtained honestly. Sixth, scientific fraud is a crime and must be investigated like any other crime. Not surprisingly, scientists are not trained to be criminal investigators and few are good at it. Seventh, I believe that laboratory directors and mentors are particularly ill-suited to carry out investigations on their own staff, as they have difficulty shedding their roles as teachers of and advocates for their trainees in order to assume the adversarial role demanded of investigators of a possible crime. Eighth, conditions in laboratories can almost always be improved. It is helpful for laboratories to prepare, adopt and adhere to specific written procedures on the conduct of experiments and the reporting of data.

Finally, I do not believe that the environment in which science is carried out is fundamentally flawed, because the overwhelming number of scientists are honest. Society at large is now grappling with the problem of developing a response to terrorism and with the question of how much of our personal freedom we must give up in an effort to minimize it. In certain respects, fraud poses a similarly vexing problem for science. The analogous question in science is how much are we prepared to change the manner in which we live our scientific lives, an integral part of which involves a sense of trust in colleagues, and how much would we lose in making this change in an effort to minimize (it is probably not possible to abolish) scientific fraud? □

Eugene Braunwald, MD is the Hersey Professor of the Theory and Practice of Medicine at the Harvard Medical School, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

Three-component isotopic heterogeneity near the Oceanographer transform, Mid-Atlantic Ridge

Steven B. Shirey*, John F. Bender† & Charles H. Langmuir‡

* Carnegie Institution of Washington, Department of Terrestrial Magnetism, 5241 Broad Branch Rd, NW, Washington, DC 20015, USA

† Department of Geography and Earth Sciences, The University of North Carolina at Charlotte, Charlotte, North Carolina 28223, USA

‡ Lamont-Doherty Geological Observatory of Columbia University, Palisades, New York 10964, USA

Tholeiitic basalts collected over an area of less than 100 km² near the Oceanographer transform fault have isotopic compositions interpreted as mixing between three chemically distinct mantle sources. Of the two enriched sources, one has an isotopic composition not previously observed in basalts from the North Atlantic. Its characteristics suggest a model whereby portions of the mantle currently beneath 35° N on the Mid-Atlantic Ridge contain detached fragments of old sub-continental mantle.

OCEANIC ridges and ocean islands, prominent surface features of the Earth's crust, reflect the composition of the underlying mantle from which they are derived. Thus the isotopic compositions of ocean ridge and ocean island basalts are some of the main data used to constrain the distribution and origin of chemical heterogeneity in the upper mantle¹⁻¹⁰. In general, two classes of models have been used to explain the origin of the observed heterogeneity on ocean ridges. Schilling and coworkers¹¹⁻¹⁷ have argued for a well-mixed, depleted upper mantle which receives injections of enriched material from discrete plumes rising from the lower mantle. These plumes, or 'hotspots', are recognized at the surface by the presence of ocean islands or very large seamounts. In Schilling's model, which he has called the 'hotspot source - ridge sink' model, the width of the chemical anomaly along the ridge is inversely proportional to the distance of the hotspot from the ridge axis. Localized 'spikes' of enrichment along the ridge are caused by hotspots on the ridge or nearby but off-axis. In contrast to this emphasis on hotspots, others^{7,8,10,18-21} favour dispersed heterogeneities of variable shape and size passively embedded in a continually remixed convecting mantle. In this model, enrichment spikes are caused by random sampling of one of the dispersed heterogeneities, with no necessary relationship to hotspots.

The evidence for the former model is based on large-scale sampling of the mid-ocean ridge at 75-100 km spacing in the vicinity of ocean islands. The evidence for the latter model is based on much smaller-scale studies, such as those near the East Pacific Rise, where relatively enriched material can exist without prominent bathymetric anomalies or large hotspots^{10,20,22}. Previously, there have been few small-scale studies using multiple radiogenic isotopes (Pb, Sr and Nd) of enriched ridge areas that could either be associated with off-axis hotspots or be sampling dispersed, enriched heterogeneities. Such an area

is the Oceanographer Fracture Zone (OFZ). This study investigates the local distribution of basalt types in the OFZ, defines the spatial (and thus temporal) extent of small-scale heterogeneity and considers the heterogeneity in the context of the two competing models. Initial isotopic and trace element results of the study are reported here; complete geochemical data will be presented elsewhere.

The Oceanographer transform fault, located at 35° N on the Mid-Atlantic Ridge, offsets the ridge right-laterally by 130 km (Fig. 1a, b). The ridge at this latitude lies within the regional geochemical gradient discovered by Schilling¹¹, which extends from the Azores platform at 39° N (Ce/Sm_N > 2.0) to the Hayes fracture zone near 34° N (Ce/Sm_N < 0.8)²³. Mixing of material from the Azores plume with depleted upper mantle is the apparent cause of the gradient¹¹. To be consistent with this regional gradient, samples from the OFZ area should have flat rare-earth-element patterns, with Ce/Sm_N of ~1.0. Superimposed on the regional gradient is a 'spike' of geochemical enrichment defined before this work by samples from two dredges, one north of the OFZ and one slightly off-axis to the south of the OFZ, which had Ce/Sm_N > 2.0 and ⁸⁷Sr/⁸⁶Sr > 0.70320^{13,14}. Rideout and Schilling²⁴ suggested that this spike was caused by the hotspot that gave rise to the New England seamounts (NES) and that the hotspot signature might also be present in drill holes on 18 and 35-Ma-old crust drilled by Leg 82 of the Deep Sea Drilling Project (DSDP)²⁵. One difficulty in interpreting small-scale heterogeneity with the existing data was the ≥100 km separation between the ridge-axis dredges and the drill holes on older crust.

Results

In this investigation we have undertaken a comprehensive geochemical study of samples from more than 60 stations in the

Fig. 1 Sample locations in the OFZ area. *a*, Numbered circles correspond to DSDP drill sites. Holes 332-335 are from Leg 37 and holes 560-564 are from Leg 82. Inset, the location of ALVIN dive 1018 and dredge RD8, both located on the south side of the transform. *b*, ALVIN dive tracks and dredge hauls on scarp northwest of the ridge-transform intersection. The double line represents the ridge axis and the straight broken line represents the transform. The outlined area encircles group A samples only.

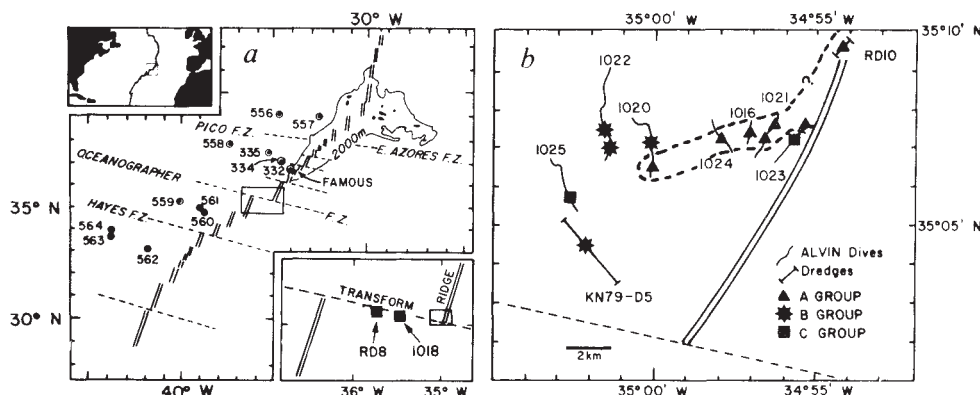


Table 1 Major, trace element and isotopic data

Group	Sample		$\frac{^{87}\text{Sr}}{^{86}\text{Sr}}$	$\frac{^{143}\text{Nd}}{^{144}\text{Nd}}$	$\frac{^{206}\text{Pb}}{^{204}\text{Pb}}$	$\frac{^{207}\text{Pb}}{^{204}\text{Pb}}$	$\frac{^{208}\text{Pb}}{^{204}\text{Pb}}$	Mg no.	$\frac{\text{K}_2\text{O}}{\text{TiO}_2}$	$\frac{\text{Ce}}{\text{Sm}_N}$	$\frac{\text{Ce}}{\text{Yb}_N}$	$\frac{\text{Nd}}{\text{Sm}_N}$	$\frac{\text{Ba}}{\text{Y}}$	$\frac{\text{Nd}}{\text{Y}}$
A	1016-1	wr	0.703411 $\pm$ 77	0.512882 $\pm$ 20	18.747	15.512	38.61	0.64	0.475	2.00	2.87	1.37	6.26	0.519
	1016-1	wrt†			18.742	15.508	38.60							
	1020-1-2	wr	0.703479 $\pm$ 20	0.512872 $\pm$ 25	18.769	15.503	38.55	0.69	0.424	2.07	2.90	1.40	5.74	0.505
	1021-2	wr	0.703512 $\pm$ 29	0.512925 $\pm$ 15§	18.765	15.491	38.54	0.59	0.520	2.24	3.72	1.47	7.50	0.663
	1021-2	gl	0.703454 $\pm$ 51	0.512890 $\pm$ 41	18.821	15.495	38.51	0.59	0.551				7.27	
	1021-4	gl	0.703515 $\pm$ 31	0.512850 $\pm$ 20	18.616	15.493	38.45	0.69	0.423	2.10	3.07	1.36	8.28	0.661
	1023-6-1	wr	0.703399 $\pm$ 20	0.512903 $\pm$ 26	18.966	15.499	38.71	0.66	0.542	2.29	3.58	1.47	7.56	0.607
	1024-1	wr	0.703460 $\pm$ 20‡	0.512860 $\pm$ 23	18.820	15.502	38.62	0.70	0.572	2.33	4.24	1.54	9.14	0.713
	1024-1	wrt†	(0.703651 $\pm$ 27)	0.512896 $\pm$ 19§	18.819	15.512	38.68							
	RD10-P4	gl	0.703500 $\pm$ 20	0.512871 $\pm$ 26	18.583	15.470	38.41	0.61	0.037	2.16	2.95	1.52	5.64	0.599
B	1020-2-1	wr	0.703242 $\pm$ 47	0.512995 $\pm$ 24	19.210	15.575	38.96	0.59	0.452	1.88	2.83	1.29	5.88	0.520
	1020-2-1	wrt†			19.186	15.560	38.92							
	1022-6	wr	0.703188 $\pm$ 20‡	0.512990 $\pm$ 26	19.538	15.592	39.17	0.68	0.484	2.47	3.53	1.47	8.80	0.560
	1022-6	wrt†	(0.703497 $\pm$ 31)		19.558	15.626	39.27							
	1022-4-2	wr	0.703073 $\pm$ 20	0.513003 $\pm$ 46	19.489	15.574	39.08	0.66	0.391	1.90	2.73	1.31	5.58	0.519
	KN79-3-5M	wr			19.452	15.562	40.00	0.62	0.285	1.84	2.61	1.33	4.30	0.536
C	1018-1-2	wr	0.703056 $\pm$ 20‡	0.513161 $\pm$ 32	18.568	15.518	38.31	0.62	0.267	1.08	1.05	1.05	2.52	0.290
	1018-1-2	wrt†		0.513122 $\pm$ 12§										
	1023-4	wr	0.703329 $\pm$ 60	0.513095 $\pm$ 14§	18.639	15.501	38.31	0.71	0.244	1.45	1.59	1.15	2.84	0.328
	1025-1	wr	0.703257 $\pm$ 35	0.513038 $\pm$ 14	18.989	15.559	38.66	0.69	0.290	1.38	1.99	1.22	2.87	0.447
	1025-1	wrt†		0.513005 $\pm$ 22§										
	1025-1	gl	0.703093 $\pm$ 31	0.512988 $\pm$ 24	18.967	15.527	38.57	0.69						
	RD8 P12	wr	0.703210 $\pm$ 43	0.513084 $\pm$ 16§	18.463	15.552	38.18	0.66	0.293	1.41	1.82	1.14	2.56	0.385
	RD8 P22	wr	0.703187 $\pm$ 28	0.513136 $\pm$ 15§	18.267	15.537	38.04	0.52	0.255	1.00	1.14	1.02	1.67	0.350

Latitudes, longitudes and depths of samples are given in ref. 27. Samples are designated as follows: wr, whole rocks, gl, glasses, wr†, replicate determination of whole rocks on separate dissolutions. $^{87}\text{Sr}/^{86}\text{Sr}$ ratios are fractionation corrected to $^{86}\text{Sr}/^{88}\text{Sr} = 0.11940$ and reported relative to an $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of 0.70800 for Eimer and Amend SrCO_3 . Strontium isotope determinations reported with ‡ have been leached in warm 2M HCl to remove seawater Sr. The values reported in parentheses directly under these leaches were determined on unleached whole rocks. All Nd and Pb isotopic compositions were measured on unleached samples. $^{143}\text{Nd}/^{144}\text{Nd}$ ratios are reported relative to a value of 0.511860 for the UCSD Nd_2O_3 Nd standard and have been fractionation-corrected to $^{146}\text{Nd}/^{144}\text{Nd} = 0.72190$. $^{143}\text{Nd}/^{144}\text{Nd}$ ratios measured at the State University of New York, Stony Brook§, are reported in the same manner. In three cases, Stony Brook analyses have been replicated on separate dissolutions to within analytical uncertainty at DTM (Department of Terrestrial Magnetism), showing good agreement between the two laboratories. Pb isotope ratios have been determined on two separate machines at DTM and have been corrected for mass fractionation to NBS 981 on each machine using the values given in ref. 99. One Pb analysis each for 1016-1 and 1020-2-1 (separate dissolutions) was done on each machine and demonstrates good agreement between the two mass spectrometers. The 2σ of the population for multiple determinations of NBS 981 gives a Pb analytical uncertainty of $0.05\% \text{ AMU}^{-1}$. In-run precision is at least an order of magnitude better and thus not quoted. Total procedural blanks obtained during the course of this study for 200 mg size samples were 700–1,000 pg for Pb, 90 pg for Nd and 300 pg for Sr. Rare-earth element data were determined at Stony Brook by isotope dilution; for analytical details and precision see Bender *et al.*²⁹. Other major and trace element data were obtained by DC plasma spectrometry at Lamont-Doherty. Mg nos have been calculated assuming Fe as 0.85 Fe^{2+} .

OFZ area. A more thorough investigation of heterogeneity has been made possible here because of the ALVIN submersible study of Fox *et al.*²⁶ and dredges from other cruises. All the samples were collected north of the transform within 10 km of the ridge (Fig. 1b)²⁷ with the exception of three samples collected on the south side of the transform ~50–60 km west of the ridge axis (Fig. 1a). The samples are mostly phyric whole-rocks showing little alteration in thin section. Glassy samples are rare but glass was analysed when present. Table 1 shows that some of the unleached whole-rock samples have a higher $^{87}\text{Sr}/^{86}\text{Sr}$ than leached whole-rocks or glasses. Because of the scarcity of glasses, we draw our petrological interpretations largely from the Pb and Nd isotope data although they are supported by the Sr isotope data.

The isotopic data and some major element and trace element ratios are summarized in Table 1 and Figs 2a–e and 3a, b. All samples are tholeiitic basalts. Based on trace element ratios, none of the samples are 'normal' depleted ocean-ridge basalts, that is, none have $\text{Ce}/\text{Sm}_N < 1$ and $\text{Ba}/\text{Y} < 1$ (see Fig. 2). The most depleted samples from the region have slightly enriched chemistry with Ce/Sm_N of 1.0–1.5 and Ba/Y of 1.6–3.0, in accord with the regional gradient extending south from the Azores¹¹. The enrichment spike inferred from two previous dredge hauls¹⁴ is strikingly confirmed by the new data, with two-thirds of the samples analysed having chemistry enriched over the regional gradient and about half of the samples with Ce/Sm_N as high as the most enriched basalts ever recovered from the global system of ocean ridges.

The trace element data might be interpreted as evidence for a single enriched source feeding the depleted mantle beneath the ridge in the OFZ area. The isotopic data, however, show that the trace element enrichment is due to two distinct enriched

sources. The two enriched groups of samples have been designated groups A and B with the most depleted group as group C. The samples in groups B and C exhibit a continuous variation in isotopic composition, which falls within the field previously defined for this region of the North Atlantic by both on- and off-axis samples (Fig. 3). We have arbitrarily divided this C–B continuum into two groups to allow comparison of the most enriched samples from it, group B, to the more homogeneous group of enriched samples, group A. Groups A and B have nearly indistinguishable concave upwards rare-earth-element patterns, in contrast to the flat rare-earth-element patterns of group C (Fig. 2a–c). In Table 1 and Figs 2d, e and 3, the group A samples are isotopically distinguished by (1) lower $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ but higher $^{87}\text{Sr}/^{86}\text{Sr}$ at similar enrichments in Ce/Sm , (2) lower $^{143}\text{Nd}/^{144}\text{Nd}$ for the same $^{206}\text{Pb}/^{204}\text{Pb}$, and (3) slightly higher $^{208}\text{Pb}/^{204}\text{Pb}$ and lower $^{207}\text{Pb}/^{204}\text{Pb}$ for the same $^{206}\text{Pb}/^{204}\text{Pb}$.

Discussion

The isotope data require melting of a source containing distinct components that have been, at some point in their history, chemically isolated for a minimum of 600 Myr. Given that the spreading age of the exposed crust in the OFZ region is only 1–5 Myr, the isotopic variability must reflect ancient mantle events. In terms of isotopic components, three distinct compositions are needed to explain the isotopic variability displayed by the OFZ data. One component has the characteristics of normal, depleted MORB (mid-ocean-ridge basalt) which is, however, not tapped as a pure end-member in this region. The second component is a high-Pb, low-Nd component similar to the isotopic composition represented by the Azores and the New

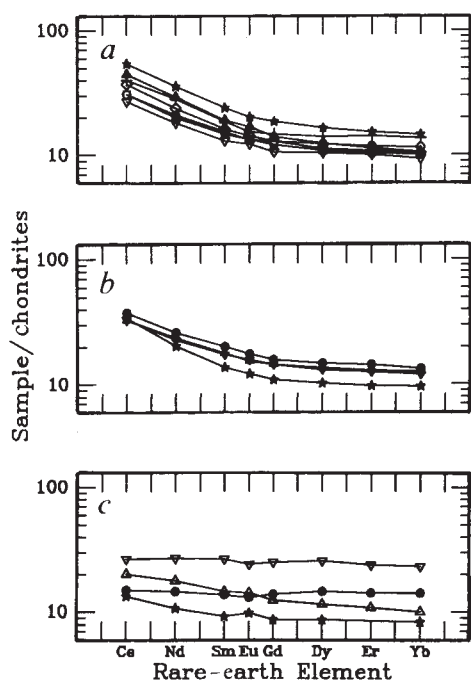
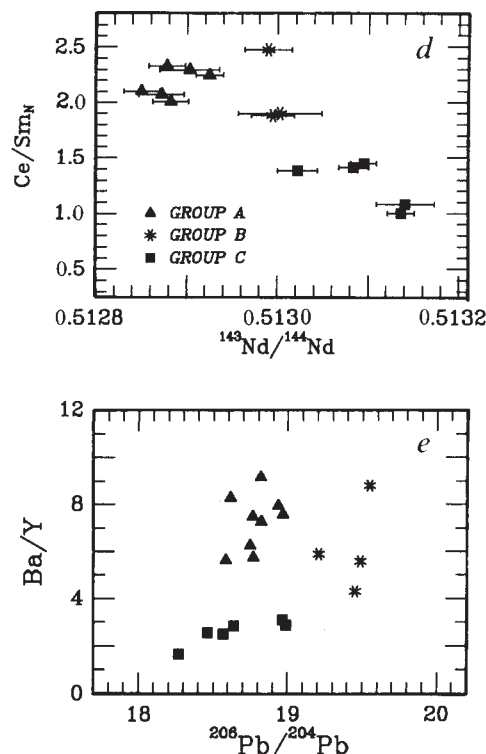


Fig. 2 *a-c*, Isotope dilution, rare-earth-element patterns for the three groups (A-C respectively) of tholeiites in the OFZ area. *d, e*, Trace element ratio plotted against isotope ratio for OFZ tholeiites ($\blacktriangle$, group A; $*$, group B; $\blacksquare$, group C). The analytical uncertainty for $^{143}\text{Nd}/^{144}\text{Nd}$ is given by error bars; the analytical uncertainty for $^{206}\text{Pb}/^{204}\text{Pb}$ is $0.05\% \text{ AMU}^{-1}$ or about $1/2$ the width of the data-points.



England seamounts (NES). The third component is a low-Pb, low-Nd component necessary to explain the group A samples (Fig. 3). This component has not been observed in the North Atlantic and, as far as is currently known, is not represented by any of the North Atlantic hotspots.

Inspection of Fig. 1*b* shows that the geographic distribution of the three chemical groups is not random. The enriched group A samples form a band which is not quite perpendicular to the ridge axis and extends over a length of about 12 km. The group B samples occur exclusively in a band approximately parallel to the ridge, about 10 km off-axis; and the group C samples occur at the westernmost and easternmost extremes of the study area. This distribution suggests that the mantle sources may not be distributed only as fine-scale (cm to m) veins but that there are discrete domains with volumes of the order of at least tens of km^3 .

The lack of spatial systematics with respect to the transform suggests that the chemical diversity is not primarily transform-related. There is no direct relation between proximity to the transform and the extent of enrichment, as might be expected from the 'transform fault effect'^{28,29}. Indeed, the two samples from stations on older crust which are closest to, but just south of, the transform are the group C samples, the least enriched chemical type. The distribution of chemical types does not appear to be affected by the transform either. The diversity in the small area north of the transform is greater than most chemical discontinuities documented across some transforms.

The chemical heterogeneity in the OFZ area is unusual in the absolute magnitude of the isotopic diversity, the type of components sampled and the presence of three chemically distinct components in a small area. In the FAMOUS region at 36°N on the Mid-Atlantic Ridge, the isotopic compositions of the basalts were found to be homogeneous^{30,31}. In other areas more diverse than FAMOUS, such as across the Kane fracture zone near 24°N on the Mid-Atlantic Ridge³² and just north of the Siqueiros fracture zone on the East Pacific Rise³³, the magnitude of isotopic variability is less and the length scale over which it is developed is greater than in the OFZ area. The one small piece of ocean crust that exhibits a similar range of variability is seamount 6 near the East Pacific Rise^{10,20}. Compared with data from this seamount, the OFZ data show a similar range in

Nd isotopic composition and a slightly smaller range in light rare-earth-element enrichment. Both seamount 6 and the OFZ area lie on the upper bounds of the oceanic mantle correlation between total isotopic variability and length scale³⁴, so neither is unique in a global sense. The OFZ area, however, does have the first-documented existence of three different chemical components in a small area. Moreover, this three-component mantle heterogeneity observed in the OFZ area does not appear to involve exactly the same end-members as the three components^{35,36} or multiple components^{7,37} proposed for other MORB and ocean island data sets. This is because of the unradiogenic $^{143}\text{Nd}/^{144}\text{Nd}$, $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ of the group A samples, coupled with their high $^{87}\text{Sr}/^{86}\text{Sr}$.

North Atlantic mantle heterogeneity. The group C to B isotopic variation in the OFZ area is similar to the regional trends from a larger area of the crust generated at the ridge over the last 35 Myr and by the two nearest hotspots active since the last 60–80 Myr (for instance Azores and NES). The unique isotopic composition of the group A samples does not fit the regional pattern and will be considered separately below. Most North Atlantic MORBs exhibit a systematic, correlated variation in Pb, Sr and Nd isotopic composition^{37–42}. The group B and C tholeiites also show this correlation, being encompassed by the MORB field including DSDP Legs 37^{43,44} and 82^{24,45,46} and the FAMOUS area on Pb–Nd, Pb–Pb isotope diagrams (Fig. 3*a, b*). Chemically, we refer to this trend as the North Atlantic MORB mantle. The isotopic composition of North Atlantic oceanic rocks is extended along the North Atlantic MORB mantle trend to more extreme values by the two hotspots in the area (Fig. 3*a, b*).

In relation to the hotspot source – ridge sink model for this region of the North Atlantic^{11,15,24,44}, the lack of normal depleted MORBs in the OFZ area has been explained by the down-ridge geochemical gradient away from the Azores plume^{14,15}, whereas the presence of an anomalous enrichment spike above the gradient has been attributed to the NES hotspot^{15,24}. The new data on the group B samples are in accord with this model because their sources are enriched in components isotopically similar to the off-axis NES hotspot. However, the isotopic composition of group B is not a particularly good constraint on the importance of the NES hotspot, because most samples from the

North Atlantic fall in the C-B trend, and the group B enrichment could be caused by a component from the Azores, the NES hotspot or some unknown source similar to that which caused the enrichment in hole 558 of Leg 82^{24,45,46}. Group B cannot be uniquely connected to the NES hotspot and the role of the NES hotspot in the OFZ enrichment can only be inferred from reconstructions of past positions of plates and hotspots^{24,47}.

Thus for the group C-B trend within the OFZ area, there may be three geographically distinct components, but there are only two chemically distinct components. The role of the Azores as a hotspot source for the OFZ area is implied by the coincidence in Ce/Sm and ⁸⁷Sr/⁸⁶Sr of the group C samples with the down-ridge geochemical gradient¹¹. The role of the NES hotspot in group B enrichment is based on proximity, isotopic composition and the idea that it would not be plausible for the hotspot producing a smooth down-ridge gradient to produce a sharp down-ridge spike. Both of these enriched components are mixed with depleted MORB; hence three geographically distinct components are present. However, only two chemically distinct reservoirs can be resolved with the isotope data because the two separate hotspots (Azores and possibly NES) and the group B samples fall along the same isotopic trends, which extend towards a high ²⁰⁶Pb/²⁰⁴Pb and ²⁰⁸Pb/²⁰⁴Pb, low ¹⁴³Nd/¹⁴⁴Nd component.

The isotopic composition of the group A samples cannot be explained by simple mixing of the mantle source for groups C and B with any of the enriched mantle sources present in the North Atlantic. The group A samples are distinct from the groups B and C mixing curves (Fig. 4a-c) and are characterized by Pb and Nd isotopic compositions that lie significantly off the MORB-Leg 82-Azores-NES isotope trend (Fig. 3a). Even mixing of MORB with a high ²⁰⁶Pb/²⁰⁴Pb, low ¹⁴³Nd/¹⁴⁴Nd component such as Sao Miguel, cannot explain the group A samples. The seven group A samples show negative correlations in Pb-Sr isotopes, and positive correlations for Pb-Nd isotopes. Compared with many enriched MORBs, they have Pb isotopic compositions that are low in ²⁰⁶Pb/²⁰⁴Pb, ²⁰⁷Pb/²⁰⁴Pb and ²⁰⁸Pb/²⁰⁴Pb. These isotopic characteristics do not correspond with any currently observed North Atlantic hotspots. Without an appropriate off-ridge hotspot with group A chemical characteristics, the hotspot source-ridge sink hypothesis seems a tenuous explanation for the observed enrichment. However, this model cannot be completely ruled out without more comprehensive isotopic analysis of other possible off-axis hotspots in the region.

An alternative hypothesis is that enrichment of the group A source may be due to passive enriched domains embedded in a matrix of North Atlantic MORB mantle. This random domain model for the 35° N area was dismissed by Rideout and Schilling²⁴ in favor of the hotspot source-ridge sink idea because all isotopic enrichment in the OFZ area was previously believed to be similar to the off-axis NES hotspot. One way to examine further the feasibility of the passive heterogeneity model is to attempt to identify the parent reservoir of the dispersed domain.

Enrichment of the oceanic mantle by incorporation of pelagic sediments^{6,39} during subduction can be ruled out because it would yield far too high ²⁰⁷Pb/²⁰⁴Pb to account for the isotopic composition of group A (Fig. 5). Incorporation of upper continental crustal sediments⁴⁸ can be eliminated using similar reasoning. Few of the ocean ridge or island samples so far analysed can specifically account for the enrichment seen in the group A samples (Fig. 5). Portions of the Pacific or Indian MORB reservoirs appear to have low enough ²⁰⁶Pb/²⁰⁴Pb and ²⁰⁷Pb/²⁰⁴Pb but these samples have too high ¹⁴³Nd/¹⁴⁴Nd^{34,38,40-42,49-52}. Many ocean island reservoirs have low enough ¹⁴³Nd/¹⁴⁴Nd but with the exception of Iceland (not shown) and Hawaii, all of them are like Tristan and Gough, for example, in having too high ²⁰⁷Pb/²⁰⁴Pb (Fig. 5). An Iceland-like reservoir is unlikely because the group A samples have higher ⁸⁷Sr/⁸⁶Sr and lower ¹⁴³Nd/¹⁴⁴Nd than any Icelandic rocks yet

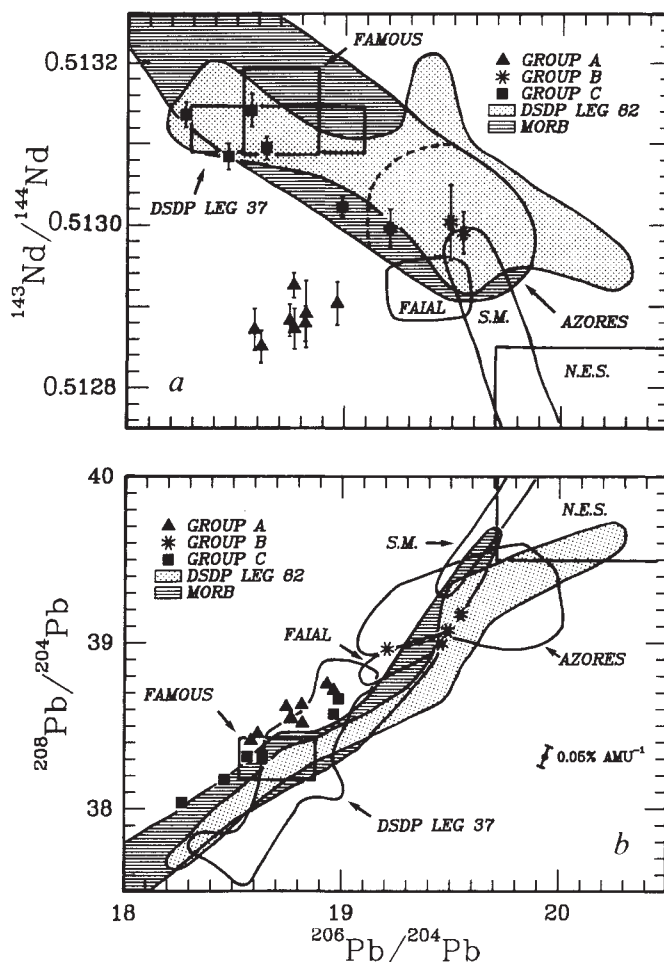


Fig. 3 Nd-Pb (a) and Pb-Pb (b) isotope compositions for Oceanographer tholeiites compared with other basalts from the nearby North Atlantic. Data sources as follows: MORB field exclusive of Indian Ocean^{38,34,39-42,61,88}, Azores Islands exclusive of Faial and Sao Miguel^{34,89}, Faial^{6,89}, Sao Miguel (S.M.)^{89,90}, DSDP Leg 82^{24,45,46}, DSDP Leg 37^{43,88}, FAMOUS^{30,31} and New England seamounts (N.E.S.)^{34,91}. Analytical uncertainty for ¹⁴³Nd/¹⁴⁴Nd is shown in a; analytical error for Pb isotopes is shown in b.

analysed³⁴, whereas a Hawaii-like reservoir (except for the Koolau volcanic series) can be discounted because of its inappropriately high ¹⁴³Nd/¹⁴⁴Nd for its ²⁰⁶Pb/²⁰⁴Pb. The Koolau series on Hawaii⁵³ does have an isotopic composition appropriate to explain the chemistry of the group A samples. However, the Hawaiian hotspot is the largest isolated hotspot on Earth and its chemistry is different from the currently known North Atlantic hotspots. Because there is no Hawaii-like hotspot in the North Atlantic with which to link the group A samples, it may be worthwhile to consider a randomly dispersed heterogeneity model to explain the incorporation of this unusual enriched component.

A sub-continental mantle source for OFZ heterogeneity? The sub-continental lithospheric mantle is an enriched reservoir with the chemical characteristics necessary to explain the enrichment of the group A tholeiites. The composition of the subcontinental mantle is thought to be generally depleted in basaltic components⁵⁴⁻⁵⁶ although it may become subsequently enriched by metasomatism involving melts⁵⁷. Typically, it has a highly variable trace element^{58,59} and isotopic composition⁶⁰⁻⁶². Figure 5a-c shows the isotopic composition of the more unradiogenic portions of the Archaean to Proterozoic subcontinental mantle thought to make up the lower part of the Wyoming⁶³⁻⁶⁵ and Kaapvaal cratons⁶². It should be noted that more radiogenic portions of the subcontinental mantle from these cratons have

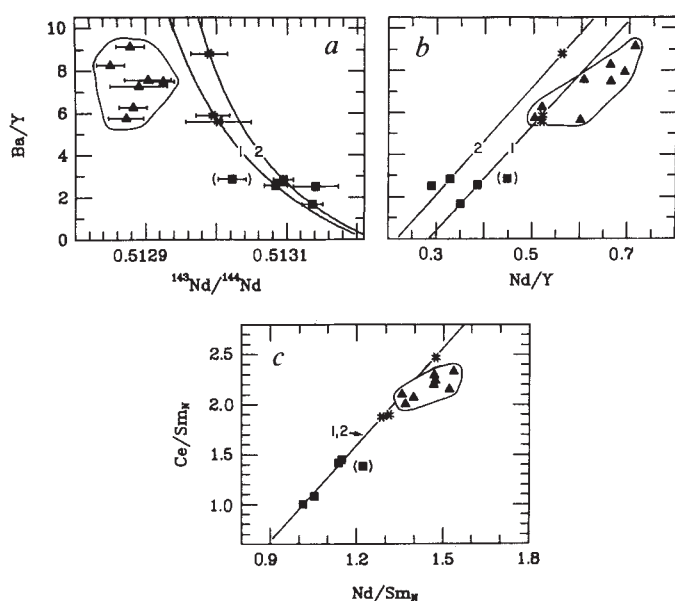


Fig. 4 Trace element ratio-isotope ratio mixing calculations; symbols as in Fig. 3. **a**, The curves through the group B and C tholeiites are calculated with the mixing equations of Langmuir⁹² using analyses for samples of similar Mg no. This method avoids fractionation effects on concentration and carries no prior assumptions about source concentration or composition. Curve 1 is calculated from samples 1022-4-2 (13.5 p.p.m. Nd, 26.0 p.p.m. Y) and P12 (10.4 p.p.m. Nd, 27.0 p.p.m. Y) and curve 2 is calculated from samples 1022-6 (12.1 p.p.m. Nd, 21.6 p.p.m. Y) and 1023-4 (6.36 p.p.m. Nd, 19.4 p.p.m. Y). Sample in parenthesis is 1025-1, which does not fall on either mixing curve. **b**, Companion plot supporting mixing of group B and group C mantle sources. Curves 1 and 2 are drawn through the samples used to plot the mixing curves. Note that multiple mixing curves are required by the Ba/Y data. In addition, the group A samples define a different slope, emphasizing the involvement of a third mantle reservoir. **c**, Companion plot using only rare-earth-element data (accompanies Fig. 2d).

been omitted from Fig. 5. This was done to clarify the figure and also because of the inability of the more radiogenic subcontinental mantle to explain the group A tholeiites. Mixing a relatively small amount (8–12%) of subcontinental mantle with a North Atlantic MORB mantle along the group C to group B trend will lower the $^{143}\text{Nd}/^{144}\text{Nd}$, $^{206}\text{Pb}/^{204}\text{Pb}$, $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ while raising the $^{87}\text{Sr}/^{86}\text{Sr}$ (Fig. 5a–c) enough to account for the composition of the group A samples. The large variability in composition of the sub-continental mantle means that there is no unique mixing curve or end-member composition. However, the family of possible mixing curves using sub-continental mantle as an end-member dictates that the North Atlantic MORB mantle end-member must be considerably enriched in NES-like components because the sub-continental lithosphere is low in both $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ (Fig. 5a, b). These features are consistent with a model for the group A mantle source as an enriched domain carrying sub-continental lithospheric mantle embedded in a matrix of North Atlantic MORB mantle. The sub-continental mantle has previously been implicated in explaining MORB heterogeneity^{51,66}, and the isotopic and trace element composition of the group A tholeiites perhaps adds to the evidence that removal and recycling of pieces of ancient sub-continental mantle into the convecting mantle could be responsible for some small-scale heterogeneities of the MORB mantle.

The distribution of heterogeneities as dispersed enriched domains is consistent with ideas about how parts of the subcontinental lithosphere might be incorporated into the convecting mantle. Based on seismic studies⁵⁴, heat flow and thermal modelling^{67,68}, geodynamic constraints⁶⁹, flood basalt

chemistry^{60,70,71}, xenolith thermobarometry⁷², mantle xenolith chemistry⁶⁰ and diamond inclusions in kimberlite⁷³, a mantle keel makes up approximately two-thirds or more of the present cratonic lithosphere. This 100–200 km-thick mantle root was frozen out of mantle convection beneath the cratonic crust during or shortly after continent stabilization and has remained physically isolated from the convecting mantle for 10^9 yr time periods. The continental lithosphere must be affected by rifting, subduction around its margins and mantle upwelling at hotspots. These processes have the potential to lead to destabilization of previously cool lithospheric mantle. Some portions of the lithospheric mantle could be removed and entrained in the flowing mantle. The subcontinental mantle, delaminated in this manner, has been suggested to cause crustal melting⁷⁴, to provide dispersed heterogeneities in the convecting mantle⁷⁵, to give rise to ocean island basalts^{61,76,77} and to contribute to certain ocean island mantle sources^{78–80}.

An estimate of the size of the enriched mantle domain in the OFZ area is given by the distribution of the group A samples (Fig. 1b). However, its true shape and original size when detached from the sub-continental lithosphere are not constrained. A heterogeneity should become thinner with age^{81,82}, but how heterogeneities survive in a convecting mantle and whether they could survive since the Archaean or Proterozoic is highly dependent on how convection is modelled. Recent calculations for whole mantle convection⁸³ show 1–2 Gyr survival times. This result differs from other studies that show the upper mantle will be somewhat mixed within 0.5–1 Gyr^{75,82}. If old heterogeneities can survive for long periods, then apparently the OFZ area shows an example of one such heterogeneity. If mantle convection is more effective in stirring the mantle towards homogeneity, then a mechanism for recycling old pieces of sub-continental mantle, such as delamination, provides an explanation of how the isotopic age of a heterogeneity could exceed the limit on its apparent physical age.

The implications of enriched domains of sub-continental lithosphere for explaining heterogeneity in the oceanic mantle are profound. Because only Precambrian materials can preserve significant primary variations in $^{235}\text{U}/^{204}\text{Pb}$, this mechanism could contribute to the long-recognized antiquity of the variation in $^{206}\text{Pb}/^{204}\text{Pb}$ against $^{207}\text{Pb}/^{204}\text{Pb}$ of the oceanic mantle^{3,5,84,85}. It could also be responsible for specific areas anomalous in $^{207}\text{Pb}/^{204}\text{Pb}$ ^{9,36}. If such dispersed, enriched heterogeneities exist in the MORB mantle, are plumes always necessary to explain enrichment spikes on the ridge? For example, such a model could provide an alternative to the hotspot source – ridge sink model for the South Atlantic¹⁵. Hawkesworth *et al.*⁷⁸ have suggested that the sub-continental mantle at the margin of the South Atlantic was responsible for the unusual Pb isotopic chemistry of the South Atlantic hotspots (Discovery, Tristan, Gough)⁵ and this signature has been shown to be faithfully reproduced in the MORB Pb isotope data¹⁷. During the opening of the present South Atlantic any detached pieces of sub-continental mantle could have been dispersed along convective flow lines. At present, enriched domains with similar isotopic composition should be sampled at the ridge as well as ocean islands, as long as they are the same spreading flow line.

Conclusions

Tholeiites from the OFZ area display much heterogeneity in trace elements and Sr, Nd and Pb isotopes: they are resolvable into three different chemical components. Tholeiites from groups C and B form a continuous enrichment trend, the group C end-member being consistent with the regional enrichment gradient from the Azores southward along the Mid-Atlantic Ridge. The group B samples of the trend are enriched in trace elements and isotopes relative to the down-ridge gradient. Several of the hotspots in this vicinity of the North Atlantic have the isotopic composition to cause this enrichment, but the New England seamount hotspot seems to be the most logical,

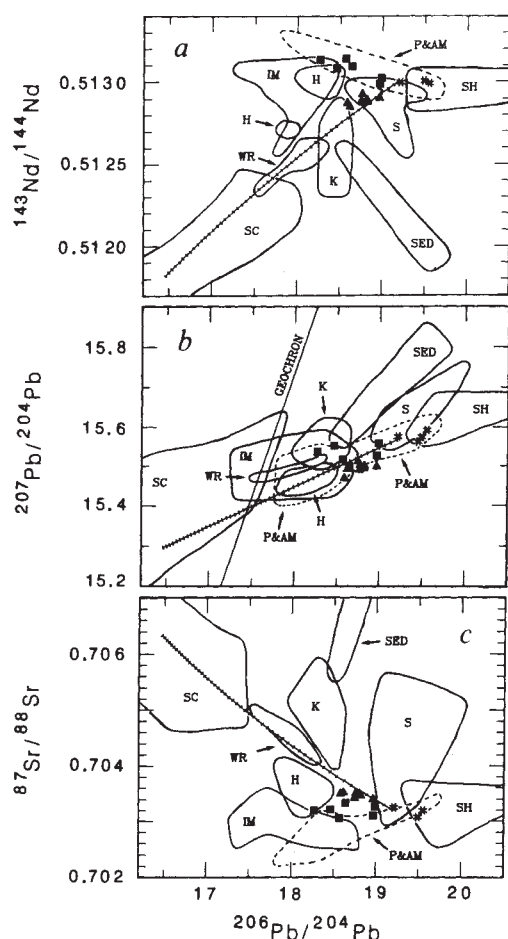


Fig. 5 Comparison of Nd-Pb (a), Pb-Pb (b) and Sr-Pb (c) data for OFZ area tholeiites (Δ , group A; $*$, group B; $\blacksquare$, group C) with the isotopic composition of proposed mantle reservoirs and sediments. Oceanic mantle reservoirs are grouped generally in accord with ref. 37, but data are taken from the literature: Saint Helena, Ascension, Australs (SH)^{5,6,37,93}; Societies, Marquesas, Sao Miguel (S)^{6,88,89,93,94}; pelagic and continental sediment (SED)^{95,96}; Kerguelen, Tristan, Gough (K)^{5,6,97}; Walvis Ridge (WR)⁶⁶; Hawaii (H)^{53,98}; Atlantic and Pacific MORB (P&AM, refs in Fig. 3); Indian MORB (IM)^{42,49-52}; and sub-continental mantle (SC). The sub-continental mantle field includes relatively unradiogenic lamproites⁶⁴, kimberlites⁶², alkaline and subalkaline volcanic rocks⁶³, high-alumina tholeiites and high-magnesian andesites⁶⁵. These samples represent the estimated isotopic composition of sub-crustal lithospheric portions of the Wyoming and Kaapval Cratons. Note that the composition of the sub-continental mantle displays a huge isotopic variability particularly in $^{207}\text{Pb}/^{204}\text{Pb}$ and it contains significant portions that have radiogenic $^{87}\text{Sr}/^{86}\text{Sr}$ but highly unradiogenic $^{143}\text{Nd}/^{144}\text{Nd}$, $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$. The Koolau volcanic series is the smaller of the two Hawaiian fields in a and makes up the less radiogenic $^{206}\text{Pb}/^{204}\text{Pb}$ end of the Hawaii field in b and c. The solid line with ticks at 2% increments displays mixing between a MORB mantle somewhat enriched in B components (30 p.p.m. Sr, 3 p.p.m. Nd, 0.4 p.p.m. Pb, $^{87}\text{Sr}/^{86}\text{Sr} = 0.70320$, $^{143}\text{Nd}/^{144}\text{Nd} = 0.51301$, $^{206}\text{Pb}/^{204}\text{Pb} = 19.20$, $^{207}\text{Pb}/^{204}\text{Pb} = 15.57$, $^{208}\text{Pb}/^{204}\text{Pb} = 38.75$) and estimates of sub-continental mantle (45 p.p.m. Sr, 4 p.p.m. Nd, 0.8 p.p.m. Pb, $^{87}\text{Sr}/^{86}\text{Sr} = 0.70629$, $^{143}\text{Nd}/^{144}\text{Nd} = 0.51184$, $^{206}\text{Pb}/^{204}\text{Pb} = 16.50$, $^{207}\text{Pb}/^{204}\text{Pb} = 15.30$, $^{208}\text{Pb}/^{204}\text{Pb} = 37.08$).

based on past plate and hotspot positions^{24,47}. A hotspot source - ridge sink model¹⁵ can account for the presence of this enriched component on the Mid-Atlantic Ridge in the vicinity of the OFZ.

The second enriched component in the OFZ area, seen in the group A tholeiites, has not previously been observed in North Atlantic MORBs. Compared to MORBs, it is characterized by relatively unradiogenic $^{206}\text{Pb}/^{204}\text{Pb}$, $^{207}\text{Pb}/^{204}\text{Pb}$, $^{208}\text{Pb}/^{204}\text{Pb}$

and $^{143}\text{Nd}/^{144}\text{Nd}$ but radiogenic $^{87}\text{Sr}/^{86}\text{Sr}$. The spatial restriction of tholeiites tapping this source to crust between 0 and 10^6 yr shows that this heterogeneity exists in the mantle on a scale larger than the scale of melting^{86,87}. Its unique isotopic composition does not correlate with the known compositions of any known North Atlantic hotspots and may reflect an isolated anomalous region entrained in the upwelling MORB mantle. The unusual chemistry of the group A samples at 35°N can be modelled by mixing 8-12% of Archaean to Proterozoic sub-continental lithospheric mantle into North Atlantic MORB mantle which is already enriched with NES-like components. Sporadic incorporation of enriched domains of sub-continental mantle may be an important mechanism to explain some small-scale or even some large-scale mantle heterogeneities that are inferred from the geochemical study of the oceanic crust.

We thank P. J. Fox, D. J. Fornari and the other members of the Oceanographer Transform Tectonic Exploration and Research Team (OTTER) for both collecting and making available the basaltic samples used in this geochemical investigation; S. DeLong of State University of New York, Albany, and D. Walker of L-DGO for providing additional sample powders; A. Biesada of L-DGO for assistance with the DC plasma analyses; G. N. Hanson of SUNY, Stony Brook, for use of mass spectrometry facilities; R. W. Carlson for discussions; and B. Dupré, P. Kenyon, J. Rubenstein, J. D. Morris and S. Esperanza for critical comments. This work was supported in part by NSF grant OCE84-00517 to J.F.B. and OCE84-11448 to C.H.L. J.F.B. also acknowledges support of the Research Foundation of the University of North Carolina at Charlotte.

Received 28 April; accepted 17 November 1986.

- Gast, P. W., Tilton, G. R. & Hedge, C. E. *Science* **145**, 1181-1185 (1964).
- Tatsumoto, M. *Science* **153**, 1094-1101 (1966).
- Tatsumoto, M. *Earth planet. Sci. Lett.* **38**, 63-87 (1978).
- Sun, S. S. & Hanson, G. N. *Geology* **3**, 297-302 (1975).
- Sun, S. S. *Phil. Trans. R. Soc. A297*, 409-445 (1980).
- White, W. M. & Hoffmann, A. W. *Nature* **296**, 821-825 (1982).
- Davies, G. F. *J. geophys. Res.* **89**, 6017-6040 (1984).
- Allègre, C. J., Hamelin, B. & Dupré, B. *Earth planet. Sci. Lett.* **71**, 71-84 (1984).
- Hart, S. R. *Nature* **309**, 753-757 (1984).
- Zindler, A., Staudigel, H. & Batiza, R. *Earth planet. Sci. Lett.* **70**, 175-195 (1984).
- Schilling, J.-G. *Earth planet. Sci. Lett.* **25**, 103-115 (1975).
- White, W. M., Schilling, J.-G. & Hart, S. R. *Nature* **263**, 659-663 (1976).
- White, W. M. & Schilling, J.-G. *Geochim. cosmochim. Acta* **42**, 1501-1516 (1978).
- Schilling, J.-G., Zajac, M., Evans, R., Johnston, T., White, W. M., Devine, J. D. & Kingsley, R. *Am. J. Sci.* **283**, 510-586 (1983).
- Schilling, J.-G. *Nature* **314**, 62-67 (1985).
- Schilling, J. G., Thompson, G., Kingsley, R. R. & Humphris, S. *Nature* **313**, 187-191 (1985).
- Hanan, B. B., Kingsley, R. H. & Schilling, J.-G. *Nature* **322**, 137-144 (1986).
- Davies, G. *Nature* **290**, 208-213 (1981).
- Morris, J. D. & Hart, S. R. *Geochim. cosmochim. Acta* **47**, 2015-2030 (1983).
- Batiza, R. & Vanko, D. *J. geophys. Res.* **89**, 11235-11260 (1984).
- Sleep, N. H. *J. geophys. Res.* **89**, 10029-10041 (1984).
- Langmuir, C. H., Bender, J. F. & Batiza, R. *Nature* **322**, 422-429 (1986).
- Bougault, H. & Treuil, M. *Nature* **286**, 209-211 (1980).
- Rideout, M. & Schilling, J.-G. *Init. Rep. DSDP* **82**, 483-496 (1985).
- Bougault, H. & Cande, S. C. (eds) *Init. Rep. DSDP* **82** (1985).
- Fox, P. J. *et al. Mar. geophys. Res.* **7**, 329-358 (1985).
- Karson, J. A. *et al. Mar. geophys. Res.* **6**, 109-141 (1984).
- Langmuir, C. H. & Bender, J. F. *Earth planet. Sci. Lett.* **69**, 107-127 (1984).
- Bender, J. F., Langmuir, C. H. & Hanson, G. N. *J. Petrol.* **25**, 213-254 (1984).
- White, W. M. *Yb. Carnegie Instn Wash.* **78**, 325-331 (1979).
- Dupré, B., Lambret, B., Rousseau, D. & Allègre, C. J. *Nature* **294**, 552-554 (1981).
- Machado, N., Ludden, J. N., Brooks, C. & Thompson, G. *Nature* **295**, 226-228 (1982).
- Carlson, R. W., McDougall, J. D. & Lugmair, G. W. *Geophys. Res. Lett.* **5**, 229-232 (1978).
- Zindler, A. & Hart, S. R. *Rev. Earth planet. Sci.* **14**, 493-571 (1986).
- Zindler, A., Jagoutz, E. & Goldstein, S. *Nature* **298**, 519-523 (1982).
- Albarede, F. *Eos* **66**, 1113 (1985).
- White, W. M. *Geology* **13**, 115-118 (1985).
- Hamelin, B., Dupré, B. & Allègre, C. J. *Earth planet. Sci. Lett.* **67**, 340-350 (1984).
- Cohen, R. & O'Nions, R. K. *Earth planet. Sci. Lett.* **61**, 73-84 (1982).
- Cohen, R., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Nature* **283**, 149-153 (1980).
- Dupré, B. & Allègre, C. J. *Nature* **286**, 17-22 (1980).
- Cohen, R. & O'Nions, R. K. *J. Petrol.* **23**, 299-324 (1982).
- Yamaguchi, M., Armstrong, R. L., Russell, R. D. & Slawson, W. F. *Init. Rep. DSDP* **37**, 613-616 (1977).
- Schilling, J. G., Kingsley, R. & Bergeron, M. *Init. Rep. DSDP* **37**, 591-597 (1977).
- Dupré, B., Gopel, C. & Bougault, H. *Init. Rep. DSDP* **82**, 497-500 (1985).
- Jenner, G. A., Hertogen, J., Sachtleben, T. & Schminke, H.-U. *Init. Rep. DSDP* **82**, 501-507 (1985).
- Duncan, R. A. *J. geophys. Res.* **89**, 9980-9990 (1984).
- Weaver, B. L., Wood, D. A., Tarney, J. & Joron, J. L. *Geology* **14**, 275-278 (1986).
- Dupré, B. & Allègre, C. J. *Nature* **303**, 142-146 (1983).
- Hamelin, B., Dupré, B. & Allègre, C. J. *Earth planet. Sci. Lett.* **76**, 288-299 (1986).
- Hamelin, B. & Allègre, C. J. *Nature* **315**, 196-199 (1985).

52. Price, R. C., Kennedy, A. K., Riggs-Sneeringer, M. & Frey, F. *Earth planet. Sci. Lett.* **78**, 379-396 (1986).
53. Stille, P., Unruh, D. M. & Tatsumoto, M. *Nature* **304**, 25-29 (1983).
54. Jordan, T. H. *Phil. Trans. R. Soc. A301*, 359-373 (1981).
55. Nixon, P. H., Rogers, N. W., Gibson, I. L. & Grey, A. A. *Rev. Earth planet. Sci.* **9**, 285-309 (1981).
56. Harte, B. in *Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 46-91 (Shiva, Nantwich, 1983).
57. Schneider, M. E. & Eggler, D. H. *Geochim. cosmochim. Acta* **50**, 711-724 (1986).
58. Frey, F. A. & Green, D. H. *Geochim. cosmochim. Acta* **68**, 1023-1059 (1974).
59. Stosch, H. G. & Seck, H. A. *Geochim. cosmochim. Acta* **44**, 457-470 (1980).
60. Hawkesworth, C. J., Erlank, A. J., Marsh, J. S., Menzies, M. A. & van Calsteren, P. in *Continental Basalts and Mantle Xenoliths* (eds Hawkesworth, C. J. & Norry, M. J.) 111-138 (Shiva, Nantwich, 1983).
61. Cohen, R., O'Nions, R. K. & Dawson, J. B. *Earth planet. Sci. Lett.* **68**, 209-220 (1984).
62. Smith, C. *Nature* **304**, 51-54 (1984).
63. Dudas, F. O., Carlson, R. W. & Eggler, D. H. *Geology* (in the press).
64. Fraser, K. J., Hawkesworth, C. J., Erlank, A. J., Mitchell, R. H. & Scott-Smith, B. H. *Earth planet. Sci. Lett.* **76**, 57-70 (1985/86).
65. Meen, J. K. thesis, Pennsylvania State Univ. (1985).
66. Richardson, S. H., Erlank, A. J., Duncan, A. R. & Reid, D. L. *Earth planet. Sci. Lett.* **59**, 327-342 (1982).
67. Vitorello, I. & Pollack, H. N. *J. geophys. Res.* **85**, 983-995 (1980).
68. Pollack, H. N. *A. Rev. Earth planet. Sci.* **10**, 459-481 (1982).
69. Oxburgh, E. R. & Parmientier, E. M. *Phil. Trans. R. Soc. A288*, 415-429 (1978).
70. Menzies, M., Leeman, W. P. & Hawkesworth, C. J. *Nature* **303**, 205-209 (1983).
71. Carlson, R. W. *Geochim. cosmochim. Acta* **48**, 2357-2372 (1984).
72. Boyd, F. R., Gurney, J. J. & Richardson, S. H. *Nature* **315**, 387-389 (1985).
73. Richardson, S. H., Gurney, J. J., Erlank, A. J. & Harris, J. W. *Nature* **310**, 198-302 (1984).
74. Bird, P. *J. geophys. Res.* **83**, 4975-4987 (1978).
75. Hoffman, N. R. A. & McKenzie, D. P. *Geophys. J. R. astr. Soc.* **82**, 163-216 (1985).
76. McKenzie, D. & O'Nions, R. K. *Nature* **301**, 229-231 (1983).
77. Allègre, C. J. & Turcotte, D. *Geophys. Res. Lett.* **12**, 207-210 (1985).
78. Hawkesworth, C. J., Mantovani, M. S. M., Taylor, P. N. & Palacz, Z. *Nature* **322**, 356-359 (1986).
79. Hart, S. R., Gerlach, D. C. & White, W. M. *Geochim. cosmochim. Acta* **50**, 1551-1557 (1986).
80. Gerlach, D. C., Hart, S. R., Morales, V. W. J. & Palacios, C. *Nature* **322**, 165-169 (1986).
81. Richter, F. M. & Ribe, N. M. *Earth planet. Sci. Lett.* **43**, 212-222 (1979).
82. Olsen, P., Yuen, D. A. & Balsiger, D. J. *Geophys. Res.* **89**, 425-436 (1984).
83. Gurnis, M. & Davies, G. F. *J. geophys. Res.* **91**, 6375-6395 (1986).
84. Chase, C. G. *Earth planet. Sci. Lett.* **52**, 277-284 (1981).
85. Church, S. E. & Tatsumoto, M. *Contr. Miner. Petr.* **53**, 253-279 (1975).
86. Bender, J. F., Shirey, S. B. & Langmuir, C. H. *Eos* **66**, 408 (1985).
87. Langmuir, C. H., Bender, J. F. & Shirey, S. B. *Eos* **66**, 408 (1985).
88. O'Nions, R. K., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **34**, 13-22 (1977).
89. Dupré, B. thesis, Univ. Paris (1983).
90. Hawkesworth, C. J., Norry, M. J., Roddick, J. C. & Vollmer, R. *Nature* **280**, 28-31 (1979).
91. Taras, B. D. & Hart, S. R. *Eos* **64**, 907 (1983).
92. Langmuir, C. H., Vocke, R. D., Hanson, G. N. & Hart, S. R. *Earth planet. Sci. Lett.* **37**, 380-392 (1978).
93. Vidal, P., Chauvel, C. & Brousse, R. *Nature* **307**, 536-538 (1984).
94. Dupré, B., Lambret, B. & Allègre, C. J. *Nature* **299**, 620-622 (1982).
95. Barreiro, B. *Geochim. cosmochim. Acta* **47**, 817-822 (1983).
96. White, W. M., Dupré, B. & Vidal, P. *Geochim. cosmochim. Acta* **49**, 1875-1886 (1985).
97. Dosso, L. & Murthy, V. R. *Earth planet. Sci. Lett.* **48**, 268-276 (1980).
98. Staudigel, H., Zindler, A., Hart, S. R., Leslie, T., Chen, C.-Y. & Claque, D. *Earth planet. Sci. Lett.* **69**, 13-29 (1984).
99. Catanzaro, E. J., Murphy, T. J., Shields, W. R. & Garner, E. L. *J. Res. natn. Bur. Stand.* **72A** (1968).

Non-tolerance and autoantibodies to a transgenic self antigen expressed in pancreatic β cells

Timothy E. Adams*, Susan Alpert & Douglas Hanahan†

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

Transgenic mice expressing simian virus 40 T antigen under control of the insulin gene regulatory region vary in their response to this protein. Each lineage is characteristically either tolerant to T antigen, or not, in which case autoantibodies arise with high frequency, and lymphocytes infiltrate and disrupt the pancreatic islets. Both non-tolerance and the autoimmune response appear to result from delayed onset of T antigen expression during β cell development.

THE way in which the immune system discriminates between self and non-self and maintains a state of tolerance to self-antigens is not fully understood. It is thought that during lymphocyte differentiation and maturation the self-reactive cells of both the cellular and humoral immune system are either physically deleted, functionally inactivated, or have their activity continuously suppressed. The complexity of the immune system, and of the experiments that have sought to describe it, has resulted in a number of theories on the mechanisms of self-tolerance¹⁻³. The failure of self-tolerance and production of high levels of autoantibodies are characteristic of a variety of chronic disease states⁴. A broad spectrum of antigenic targets are recognized, including hormones, DNA, ribonucleoprotein complexes and nuclear proteins. Autoantibody production may also be important in the pathogenesis of insulin-dependent (type I) diabetes mellitus (IDDM). In IDDM, 60-85% of newly diagnosed patients have islet-cell specific autoantibodies⁵. The appearance of islet-cell autoantibodies implies that β cells are being destroyed; this may occur up to eight years before the clinical manifestation of the disease⁶.

Recently one of us (D.H.) established several lineages of transgenic mice carrying integrated copies of a hybrid gene comprising the upstream regulatory sequences of the rat insulin

II gene linked to protein coding information for the simian virus 40 (SV40) large T antigen⁷. Immunohistological analysis of pancreatic tissue from transgenic mice showed that large T antigen was expressed in a tissue-specific manner, being found exclusively in the β cells of the islets of Langerhans. Immunochemical and immunohistological studies on a variety of other tissues have not revealed additional sites of T antigen synthesis. The expression of large T results in hyperplasia of the pancreatic islets and formation of β cell tumours. During these studies we became interested in the immunological consequences of transferring foreign genes into the mouse germ line. One might expect the immunological response to a protein produced by a stably integrated, inherited transgene to be similar to that of other self-proteins, that is, tolerance, but, surprisingly, this is not always the case. In three of the four mouse lineages transgenic for the insulin-promoted SV40 T antigen gene, a high incidence of serum autoantibodies to large T antigen is observed.

Autoantibodies to SV40 large T

Sera were isolated from randomly selected mice of various ages (six weeks to ten months) comprising control (B6D2F1/J strain) animals and representative animals of the RIR-Tag, RIP-Tag2, -Tag3 and -Tag4 transgenic lineages⁷. The presence of serum autoantibodies to large T antigen was assayed by immunoprecipitation of T antigen from extracts of biosynthetically-labelled COS cells, a transformed primate cell line constitutively expressing T antigen⁸. Immunoprecipitates were

* Present address: Dept of Preclinical Veterinary Science, University of Melbourne, Parkville, Victoria 3052, Australia.

† To whom correspondence should be addressed.

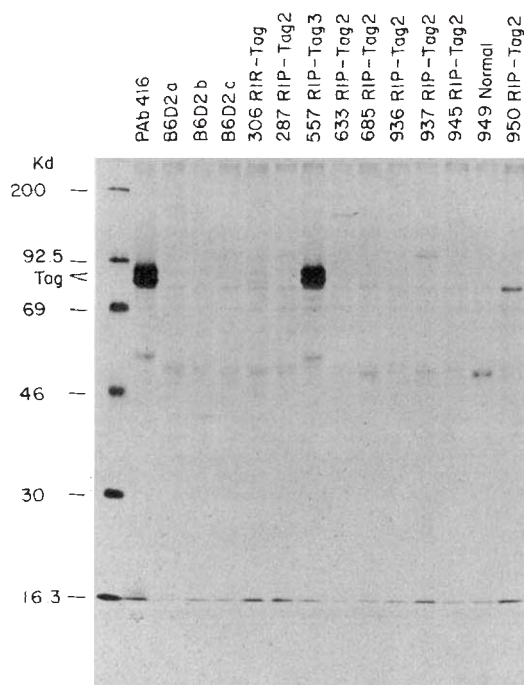


Fig. 1 Autoantibodies of SV40 large T antigen in sera from transgenic mice. Detergent-solubilized lysates of ^{35}S -methionine-labelled COS cells were immunoprecipitated with a large-T antigen-specific monoclonal antibody (PAb416)⁹, sera from B6D2 mice (tracks a-c) or with sera obtained from randomly selected mice of the RIR-Tag, RIP-Tag2 and -Tag3 transgenic lineages. Serum from an animal (949) that was negative for integrated copies of the insulin promoter-SV40 T antigen gene (a result of a mating between a normal and a transgenic mouse) was included. Immunoprecipitations were analysed by electrophoresis on 10% SDS-polyacrylamide gels and gel autoradiography.

Methods. Immunoprecipitations were performed as described²⁰, with minor modifications. Semi-confluent monolayers of COS cells⁸ were labelled 4 h with ^{35}S -methionine (Amersham) in Dulbecco's modified Eagle medium (minus methionine) and cell extracts prepared by lysis with 1% NP-40. Lysates were precleared overnight by incubation at 4 °C with 10 μl of normal rabbit serum and 100 μl of a 10% suspension of fixed *Staphylococcus aureus* Cowan 1 strain cells. Aliquots of lysate supernatant were reacted with 1 μl of mouse serum for 30 min at 4 °C. Rabbit anti-mouse immunoglobulin antisera (1 μl) was added and the incubation continued for 30 min. Immune complexes were collected by absorption to protein A-Sepharose beads (Pharmacia), washed, eluted and analysed on 10% SDS-polyacrylamide gels²⁰.

analysed by SDS-polyacrylamide gel electrophoresis and gel autoradiography (Fig. 1). A characteristic large-T-antigen doublet of relative molecular mass 85,000–90,000 (M_r 85–90K) is immunoprecipitated by the T antigen-specific monoclonal antibody PAb416 (ref. 9). Sera from control mice and from the RIP-Tag2 lineage did not immunoprecipitate detectable amounts of T antigen. In contrast, strongly immunoprecipitating antibodies are found in serum from 557, a RIP-Tag3 mouse. Similarly, weakly immunoprecipitating antibodies are observed in 306, a RIR-Tag animal, reflecting a low antibody titre.

The pattern of autoantibody incidence is clearer when the results from the random screen of transgenic mice are tabulated (Table 1). With one exception (a weak positive), sera obtained from RIP-Tag2 mice did not immunoprecipitate T antigen, but sera from RIP-Tag3, RIP-Tag4 and RIR-Tag mice did. The high incidence (61%) of autoantibody-positive sera from RIP-Tag3 animals compared to these from RIP-Tag4 animals (35%) is intriguing, as the relative frequencies were stable in successive generations, and are apparently characteristic of the lineages.

Table 1 Frequency of autoantibodies to SV40 large T antigen in sera from different transgenic mouse lineages

Lineage	No. of mice sampled	No. of mice autoantibody positive
RIP-Tag2	26	1* (4%)
RIP-Tag3	71	43 (61%)
RIP-Tag4	34	12 (35%)
RIR-Tag	11	3 (27%)

Serum was isolated from randomly selected transgenic mice of various ages and screened by immunoprecipitation for the presence of antibodies against SV40 large T antigen, as described (Fig. 1). Twenty of the RIP-Tag3 and -Tag4 mice were H-2 typed, in an assay of complement-mediated lysis of splenocytes using H-2 type specific monoclonal antibodies (a gift of U. Hammerling). Thirteen were H-2^b/H-2^b, six were H-2^b/H-2^d, and one was H-2^d/H-2^d. There was no correlation between the H-2 haplotype and the presence of autoantibodies. This is consistent with the observed ability of both haplotypes to respond to congenic SV40 transformed cells^{21–23}, and to purified large T-antigen²¹.

* Weak positive reaction.

The appearance of T antigen autoantibodies is age related; 48% (12/25) of RIP-Tag3 mice under six months of age had detectable autoantibody responses, whereas 67% (31/46) of the mice over six months of age had serum autoantibodies against T antigen (data not shown). Autoantibodies specific for T antigen have never been observed in sera taken from normal (non-transgenic) mice.

Autoimmune diseases are often associated with specific haplotypes of the major histocompatibility complex, although in general strict causal relationships have not been established. The original RIP-Tag transgenics were F₂ hybrids between the inbred strains C57B1/6J (H-2^b) and DBA/2J (H-2^d). These founder mice were backcrossed to C57B1/6J to establish lineages. It is therefore possible that differences in the H-2 haplotypes of the different lineages are affecting the autoimmune response. To address this, mice from each lineage were analysed for both autoantibodies and H-2 type (Table 1). No correlation between the two was observed, nor has continuous breeding altered the frequency of autoantibodies. These data argue against a strict linkage to either H-2 haplotype.

No link between autoantibodies and tumours

These lines of transgenic mice heritably develop highly vascularized tumours that are composed of β cells expressing large T antigen, raising the possibility that the appearance of autoantibodies to T antigen is a direct consequence of tumour formation. To assess this possibility, 27 mice of the RIP-Tag3 and -Tag4 lineages were killed, and their pancreases examined under a dissecting microscope for the presence of vascularized tumours. Serum was isolated and analysed for T-antigen specific antibodies. The results for six mice are shown in Fig. 2. Autoantibodies were not observed in sera from three out of five tumour-bearing mice, whereas two mice showed both autoantibodies and tumours. The sixth mouse shown in Fig. 2, 1806, displayed a strong antibody response to large T in the absence of identifiable tumours. This has proved to be a common observation as these analyses have been extended to mice in the RIR-Tag and RIP-Tag4 lineages, as well as to other RIP-Tag3 mice. In the study of 27 RIP-Tag3 and -Tag 4 mice, 17 were autoantibody positive, of which nine had tumours and eight did not. Similarly, four of the ten autoantibody-negative mice had tumours. Of the mice under six months old, four out of seven were autoantibody positive, and none of the seven had tumours. Thus it seems that the humoral autoimmune response can precede tumour formation, and that the induction of autoantibodies against T antigen is not an obligatory consequence of the presence of β cell tumours.

Self-tolerance assays

A number of phenotypic differences between these transgenic lineages have been observed. Mice of the RIP-Tag2 lineage develop multiple β -cell tumours and die between ten and sixteen weeks of age. In contrast, mice of the RIR-Tag, RIP-Tag3 and -Tag4 lines develop tumours more slowly, and live for six to twelve months. As mentioned previously, there appears to be an increased incidence of T-antigen-specific autoantibodies in sera taken from older transgenic mice. The absence of autoantibodies in the sera of RIP-Tag2 mice might therefore simply reflect the premature death associated with this lineage. To examine this question, four normal mice and four transgenic mice from each lineage, all six to eight weeks of age, were immunized with immunoaffinity-purified large T antigen, and their primary and secondary antibody responses were analysed. The control mice, together with all individuals in the RIR-Tag, RIP-Tag3 and -Tag4 lineages, all mounted a readily detectable primary antibody response as a result of a single immunization with T antigen (Fig. 3). In contrast, only one individual (RIP-Tag2c) of the four RIP-Tag2 mice immunized displayed a detectable, weak, primary response.

The secondary immune responses of the different lineages to a subsequent immunization with T antigen is even more striking (Fig. 3). Strong secondary immune responses to T antigen were found in the sera from all the RIR-Tag, RIP-Tag3 and Tag4 mice, comparable to that observed in normal mice. In comparison, two RIP-Tag2 mice (c,d) produced weakly immunoprecipitating antibodies, while two others (RIP-Tag2a,b) did not respond. Thus it would appear that mice of the RIP-Tag2 lineage are tolerant to the SV40 large T antigen, whereas the other lineages are non-tolerant.

Pancreas histopathology

The frequent appearance of circulating autoantibodies to a protein which is expressed at detectable levels only in pancreatic β cells raises the possibility that this observation is associated with an immune response against the β cells. Four mice bearing autoantibodies have been analysed by treating cryostat sections

of pancreas with histochemical stains such as haematoxylin and eosin, or Giemsa. These analyses have revealed frequent abnormalities, including extensive vascularization, fibrosis, the presence of large and numerous lymph nodes, and clear indications of lymphocyte infiltration of the islets of Langerhans, with varying degrees of obvious disruption (Fig. 4). The islets contain small, atypical cells in the region normally composed primarily of β cells. These cells show histochemical staining patterns characteristic of lymphocytes. Similar cells are present in adjacent enlarged blood vessels or lymph nodes.

A spectrum of islet morphologies, ranging from virtually complete disruption to relative normality, is observed in the two autoimmune pancreases shown in Fig. 4. Multiple sections were taken through each pancreas so as to represent the entire tissue. At least 30 islets were evaluated for lymphocyte infiltration, which was found in 55% and 60% of the islets of the two mice. Neither of these mice had vascularized tumours, either visibly or histologically. The observed insulinitis is, therefore, almost certainly a response against the β cells in the islets of Langerhans. A mouse from the RIP-Tag3 lineage that was immunized with exogenous large T protein also showed lymphocyte infiltration in 80% of the pancreatic islets (Fig. 4). This suggests that the immune system is capable of both locating and responding to cells that synthesize large T, regardless of the route by which the immune response to T antigen was first initiated.

In tests of more than twenty mice from the tolerant RIP-Tag2 family we found no evidence of the abnormalities described

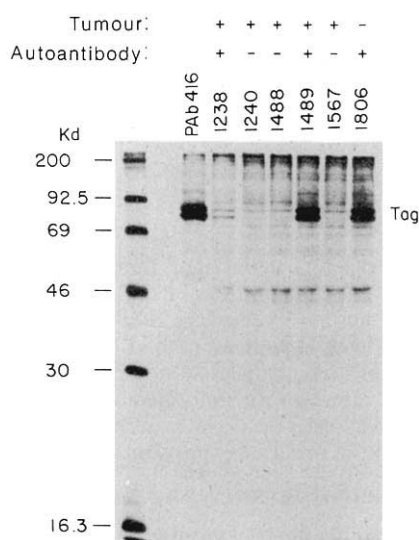


Fig. 2 Serum analysis of tumour and non-tumour bearing transgenic mice. The presence of autoantibodies to large T antigen in the sera of several tumour-bearing mice (1238, 1240, 1488, 1489, 1567) and one non tumour-bearing mouse (1806) of the RIP-Tag3 lineage was analysed by immunoprecipitation from 35 S-methionine-labelled COS cells as described (Fig. 1). Included in the analysis was the T-antigen-specific monoclonal antibody PAb416. The presence of pancreatic tumours was determined by microscopic analysis of the pancreas after killing individual mice. Blood for serum isolation was collected concurrently.

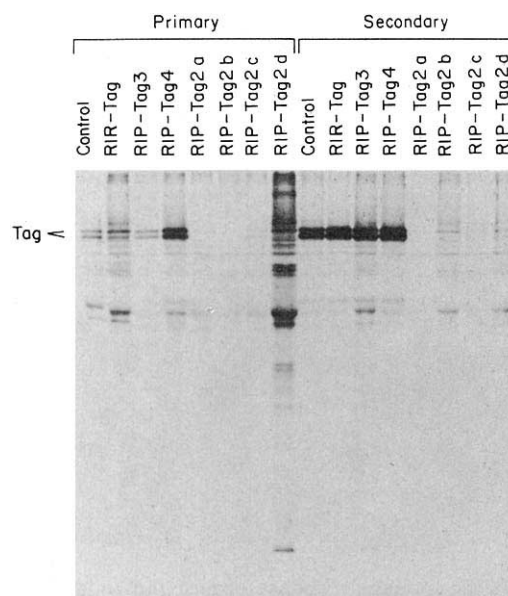


Fig. 3 Primary and secondary antibody responses of transgenic mice immunized with purified SV40 large T antigen. Depicted in the figure are the primary and secondary antibody responses of random individuals representing the non-transgenic control, RIR-Tag, RIP-Tag3 and RIP-Tag4 lineages, as well as the responses of four individual mice of the RIP-Tag2 (a-d) lineage. Four mice from each lineage were analysed (including the control). A typical result for each responding type is shown.

Methods. Immunoaffinity purified SV40 large T antigen (a gift from Y. Gluzman) was used as the immunogen. Mice (6-8 weeks old) of the RIP-Tag2, -Tag3, -Tag4 and RIR-Tag transgenic lineages were bled and screened for the presence of serum autoantibodies to T antigen. The mice were immunized (day 0) intraperitoneally (i.p.) with 10 μ g of purified T antigen in complete Freund's adjuvant in a final volume of 0.2 ml. Fourteen days later the animals were boosted with a further 10 μ g of antigen (in incomplete Freund's adjuvant) administered i.p. Mice were tail-bled on days 7 (primary response) and 21 (secondary response), serum isolated, and the antibody response to T antigen analysed by immunoprecipitation as described (Fig. 1). Included in the study as normal controls were mice of the B6D2F1/J strain.

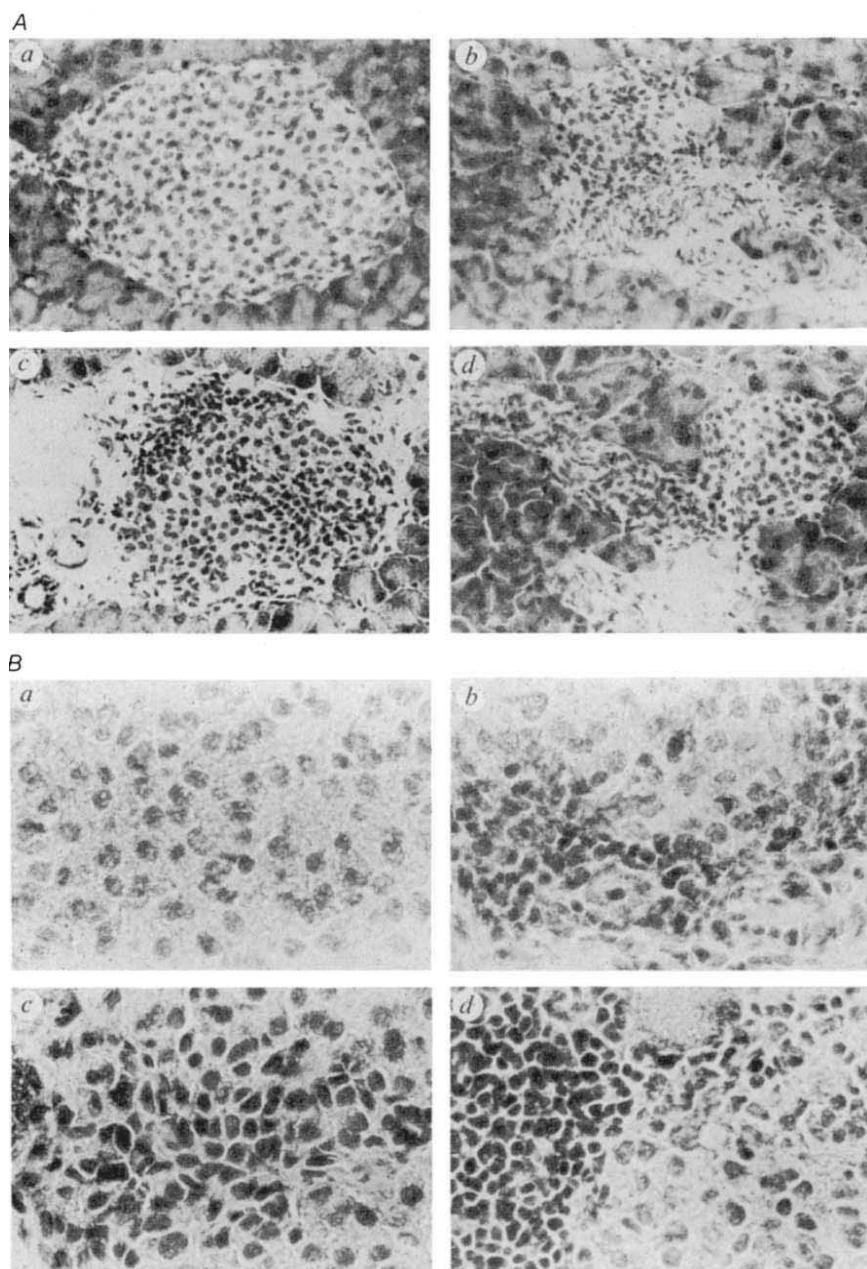


Fig. 4 Pancreas histopathology in mice producing antibodies against T antigen. Two mice bearing autoantibodies (2365 and 3217), and one with induced antibodies (1812) against SV40 large T antigen were analysed by cryostat sectioning of the pancreas followed by histochemical staining with hematoxylin and eosin. Mouse 2365, a RIP1-Tag3 mouse, was 8½ months old at the time of this analysis. The pancreas had β cell hyperplasia but no distinct tumours (either visibly or histologically). Mouse 3217, a RIP-Tag4 mouse, was 6 months old and did not have significant β cell hyperplasia or solid tumours, although enlarged lymph nodes were present. Mouse 1812 was immunized with 20 μ g of purified SV40 large T protein at 6 weeks of age, and analysed at 11½ months of age. No tumours or hyperplasia were noted in mouse 1812. **A**, Characteristic morphology of pancreatic islets in these mice: *a*, normal, non-transgenic mouse islet; *b*, islet from mouse 1812, who was immunized with Tag; *c*, islet from mouse 2365; *d*, islet from mouse 3217, showing an adjacent aggregation of lymphocytes (other islets in this pancreas had large holes as seen in *c*). All magnifications were 160 $\times$. **B**, High magnification (400 $\times$) photos showing the interior regions of pancreatic islets: *a*, normal islet full of β cells; *b*, islet from mouse 1812, immunized with exogenous large T at a different location (i.p.); *c*, islet from mouse 2365; *d*, segment of a lymph node found within a hyperplastic islet in the pancreas of mouse 2365.

Methods. Pancreases were removed in phosphate buffered saline and infiltrated with 30% sucrose. The tissues were frozen in Lipshaw mounting media and sectioned on a cryostat microtome at 10 μ m. The sections were stained with a standard histochemical procedure for haematoxylin and eosin (Sigma).

above. Thus the immune response against the endocrine pancreas occurs only in lines of mice that are not tolerant to this transgenic β cell antigen.

Ontogeny of T antigen expression

The establishment of tolerance to self must require presentation (and hence expression) of self antigens, probably during late embryogenesis and early neonatal life. Mice from each of the families of transgenic mice harboring insulin-T antigen genes have been examined for expression of large T antigen by immunohistochemistry, using antibodies which specifically recognize this protein. An analysis of neonatal and adult mice from the tolerant RIP-Tag2 lineage, from the non-tolerant RIP-Tag3 and -Tag4 lineages and of normal non-transgenic mice is shown in Fig. 5. Large T is clearly expressed at high levels in RIP-Tag2 mice, both neonatally and in adults. Similar analyses of 22 mice in this family from the last week of embryogenesis throughout adulthood reveals that T antigen is expressed at high levels in β cells at all times.

In contrast, the non-tolerant lines do not show T antigen expression in neonatal mice, and even by ten weeks of age

expression is heterogenous in the islets. Immunostained sections of 40 islets from four mice (10 weeks old) in the non-tolerant lineages showed that 0.0–3.3% of the cells in an islet had high levels of large T. This is in contrast to the RIP-Tag2 mice, where 35–100% of the cells in each islet section contain T antigen. It thus appears that, although the non-tolerant lines express large T specifically in the β cells (as do mice in the RIP-Tag2 family), there is a delay in the onset of expression.

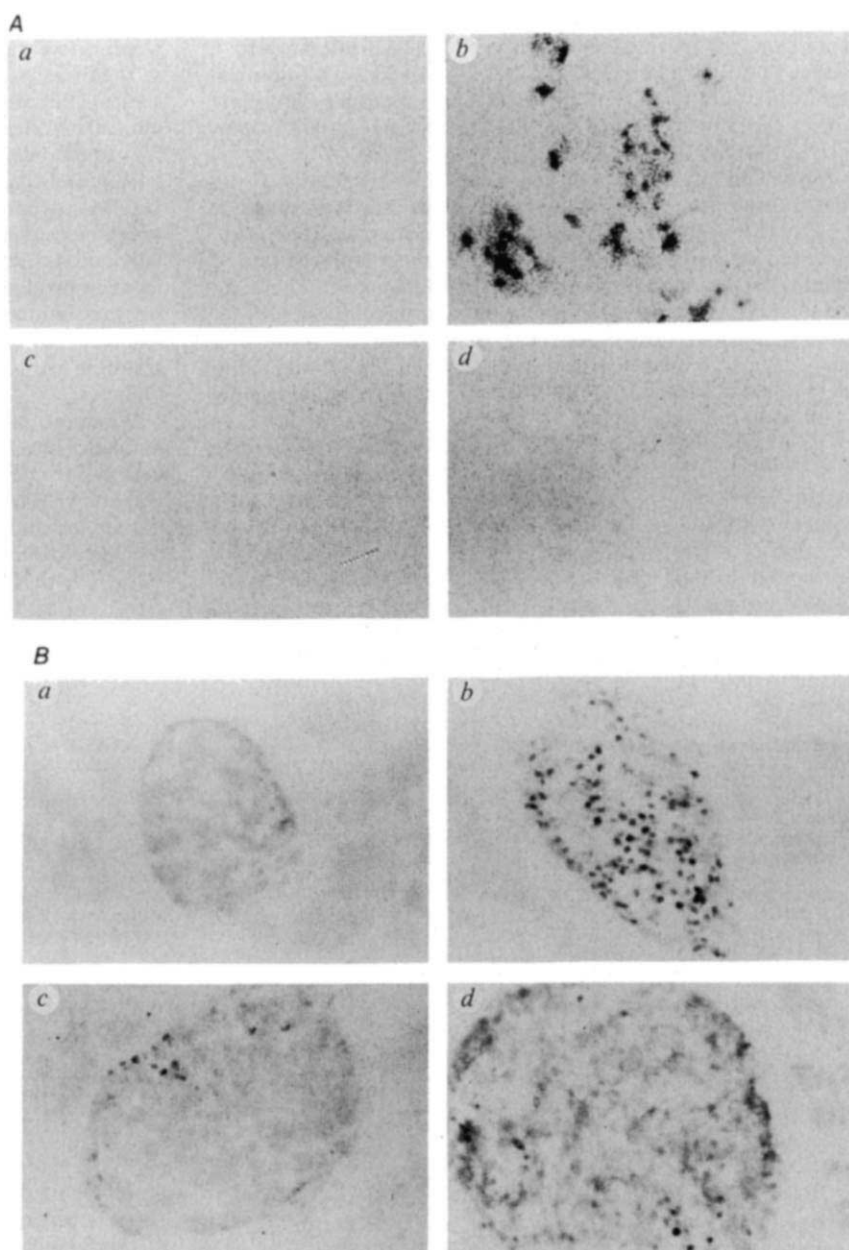
A model for self-tolerance and autoimmunity

We have shown that the delayed onset of expression of a hybrid insulin/SV40 T antigen gene in several lines of transgenic mice results in failure to establish self-tolerance and, consequently, production of antibodies specific for large T antigen. We propose that similar molecular mechanisms may be involved in the aetiology of diseases of apparent autoimmune origin, in particular insulin-dependent diabetes.

The expression of a number of intact and hybrid genes has been shown to be tissue-specific in transgenic mice¹⁰. However, the transgenes often have unusual characteristics of expression when compared to the corresponding endogenous genes, includ-

Fig. 5 Ontogeny of expression of T antigen in the pancreas. Neonatal and adult pancreases from normal mice and from RIP-Tag2, -Tag3, and -Tag4 mice were analysed for the presence of T antigen by immunostaining. **A**, analysis of neonatal mice: *a*, normal; *b*, RIP-Tag2; *c*, RIP-Tag3; *d*, RIP-Tag4. The primary polyclonal anti-Tag antibody was visualized using an ^{125}I -labelled second antibody and autoradiography. This method allows identification of very low levels of T antigen expression. Note that no expression is detectable in the RIP-Tag3 and -Tag4 lineages. The grains in the RIP-Tag2 pancreas (*b*) show characteristic nuclear localization for large T. **B**, analysis of 10-week-old mice from the same lineages, in the same order, using immunoperoxidase visualization of the antibody-antigen complex. The magnification is $200\times$ in **A**, and $160\times$ in **B**.

Methods. Neonatal mice were killed by cervical dislocation and dissected in phosphate buffered saline. The pancreas was removed into a fixative solution of 4% paraformaldehyde buffered to pH 7.4 with 0.1 M sodium phosphate buffer and postfixed for 1 h. The fixed tissues were infiltrated with 30% sucrose, mounted in Lipshaw M1 medium, sectioned in a cryostat microtome, and the sections mounted on gelatin-coated slides. Adult mice were perfused through the heart with the fixative solution, the pancreas was removed and postfixed for 1 h. Further processing was carried out as described above. The sections were washed in Tris-buffered saline (TBS, 0.1 M Tris, pH 7.5 and 0.9% NaCl), exposed first to 0.25% Triton X-100 in TBS for 15 min, then to goat serum diluted 1:30 in TBS for 30 min, and incubated overnight with a rabbit polyclonal antibody to T-antigen (a gift of David Lane). After the primary antibody treatment (1:500) the sections were treated for immunoperoxidase staining as described²⁴. The sections were washed with TBS, incubated with a 1:50 dilution of goat anti-rabbit IgG (Miles Scientific) for 30 min, rinsed with 1% goat serum in TBS, and then exposed to a 1:100 dilution of rabbit peroxidase anti-peroxidase (gift of G. Teitelman) for 30 min. Bound peroxidase was visualized by reaction with 0.25 mg ml⁻¹ of 3,3'-diaminobenzidine (DAB, Aldrich), 3 ml⁻¹ of nickel sulphate (Sigma) and 0.01% hydrogen peroxide. After the DAB reaction the sections were dehydrated through graded alcohols and mounted with Permout. For autoradiography, the sections were washed with 1% goat serum in TBS after primary antibody reaction (1:5,000), then incubated for 30 min in a 1:50 dilution of donkey anti-rabbit IgG labelled with ^{125}I (specific activity of 5–20 $\mu\text{Ci } \mu\text{g}^{-1}$ Amersham). The sections were washed, air-dried overnight and dipped in Ilford L-4 photographic emulsion (Polysciences) diluted 1:1 in water. The slides were air-dried and placed in a light-tight box containing a drying agent at 4 °C. After 3 weeks the sections were developed for 2 min at 16–17 °C in Kodak D-19 developer, rinsed in Kodak Ektaflo (diluted 1:3 in water) for 8 minutes at 4 °C, and then dehydrated through graded alcohols and mounted with Permout.



ing failure to express at all. It has been suggested that transgenes integrated into 'transcriptionally silent' regions of chromatin (for example, heterochromatin) may be incapable of expression^{10,11}. This type of nonspecific dominant repression, which might be termed a position effect, could be the basis for the differential patterns of T antigen expression in the RIP-Tag3 and -Tag4, and RIR-Tag lineages when compared to the RIP-Tag2 mice. The subsequent tissue-specific expression of T antigen in RIP-Tag3, -Tag4, and RIR-Tag mice may result from the relaxation of this non-specific repression as the cells age; eventually the β cell specific trans-acting factors are allowed access to the β cell specific regulatory region which resides in this inaccessible location in the genome.

Why are these lines of transgenic mice heritably non-tolerant to large T antigen? It is well established that neonatal mice are readily tolerized to foreign antigens^{1–3}, indicating that this stage

of development is critical for the ability of a maturing immune system to establish a mechanism of self-discrimination. In the RIP-Tag2 mice, embryonic/neonatal expression of large T antigen in the pancreas is apparently sufficient to establish T antigen-specific tolerance. As presentation of an antigen to the immune system during the period of self-learning is thought to require its circulation, it will be important to determine serum levels of T antigen shortly after birth. RIP-Tag2 mice can produce antibodies that weakly immunoprecipitate T antigen after immunization with large T protein, suggesting that the B lymphocytes capable of recognizing T antigen have not been completely depleted. This interpretation is consistent with the inducibility of B cell responses to proteins normally recognized as self, as discussed recently by Mitchison¹². An argument can thus be made that the tolerance to large T resides in another compartment of the immune system, perhaps in that of the T helper or

T suppressor cells. Because of delayed developmental expression, the mechanism which produces self-tolerance to T antigen is not effective in RIP-Tag3, -Tag4 and RIR-Tag mice. The availability of these transgenic mouse lineages will allow further experimentation to elucidate the cellular basis of tolerance to this transgenic self-antigen.

The results presented here also motivate an analysis of the time course of the autoimmune response in individual mice. It is possible that the immune response will persist in some mice, and die out in others, which may in turn affect both the rate of tumour formation and the extent of islet destruction. There are two competing processes in the autoimmune mice: β cell proliferation and β cell destruction. This may sometimes result in a balance, which could be relevant to strategies for counteracting the slow β cell destruction in type 1 diabetes by inducing slow β cell proliferation.

The observation that delayed expression of β -cell specific large T antigen results in autoantibody (and presumably cellular) responses is relevant to a number of autoimmune diseases such as insulin-dependent diabetes. Many factors may lead to the initiation of IDDM development, but it would appear that the destruction of pancreatic β -cells is immune-mediated^{5,13}. Autoantibodies to cytoplasmic¹⁴ and cell-surface¹⁵ antigens of islets are prevalent, and lymphocyte infiltration of the pancreatic islets has been documented^{16,17}. It is notable that in type I

diabetes of humans and rats, an islet cell protein of M_r 64K appears to act as a major antigenic target during the development of the disease. Autoantibodies to the 64K protein appear in the sera of humans and IDDM-susceptible rat strains well before clinical manifestation of the disease are apparent^{18,19}. It is important to establish the developmental pattern of expression of the 64K protein and other major antigenic targets in normal and IDDM-susceptible individuals. Perhaps the delayed onset of expression of a β cell specific gene, such as that encoding the 64K autoantigen, results in failure to establish self-tolerance to its gene product. The consequence of this non-tolerance might be the induction of humoral and/or cellular autoimmune responses when the 'foreign' β cell antigen eventually appears, resulting in β cell destruction and the development of diabetes.

We thank Jacek Skowronski for recent data on autoantibody incidence and for comments on the manuscript; Steinunn Baekkeskov for comments on the manuscript; Yasha Gluzman and Robert Tjian for gifts of purified T antigen; Ulrich Hammerling for advice and materials for H-2 typing; David Lane for T-antigen antiserum; Gladys Teitelman for advice on immunohistology; Seth Grant for discussions; Jean Roberts and Dave Green for art work and photography and Marilyn Goodwin for preparing the manuscript. This research was supported by a grant from Monsanto Company.

Received 12 September; accepted 19 November 1986.

1. Weigle, W. O. *Adv. Immun.* **16**, 61-123 (1973).
2. Howard, J. G. & Mitchison, N. A. *Prog. Allergy* **18**, 43-96 (1975).
3. Nossal, G. J. V. *A. Rev. Immun.* **1**, 33-62 (1983).
4. Smith, H. R. & Steinberg, A. D. *A. Rev. Immun.* **1**, 175-210 (1981).
5. Baekkeskov, S., Christie, M. & Lernmark, A. in *Immunology of Endocrine Diseases* (ed. A. McGregor) 73-88 (MTP Press, Lancaster) (1986).
6. Srikantha, S., Ganda, O. P., Eisenbarth, G. S. & Soeldner, J. S. *N. Eng. J. Med.* **308**, 322-325 (1983).
7. Hanahan, D. *Nature* **315**, 115-122 (1985).
8. Gluzman, Y. *Cell* **23**, 175-182 (1981).
9. Harlow, E., Crawford, L. V., Pim, D. C. & Williamson, N. M. *J. Virol.* **39**, 861-869 (1981).
10. Palmiter, R. & Brinster, B. A. *Ret. Genet.* **20**, 465-499 (1986).
11. Ornitz, D. M. *et al. Nature* **313**, 600-602 (1985).

12. Mitchison, N. A. *Clin. Immun. Newsletter* **6**, 1-12 (1985).
13. Rossini, A. A., Mordes, J. P. & Like, A. A. *An. Rev. Immun.* **3**, 289-320 (1985).
14. Bottazzo, G. F., Gorsuch, A. N., Dean, B. M., Cudworth, A. G. & Doniach, D. *Lancet* **i**, 668-672 (1980).
15. Lernmark, A. *et al. N. Engl. J. Med.* **299**, 375-380 (1978).
16. Geps, W. & LeCompte, P. M. *Am. J. Med.* **70**, 105-115 (1981).
17. Bottazzo, G. F. *et al. New Engl. J. Med.* **313**, 353-360 (1985).
18. Baekkeskov, S., Dynberg, T. & Lernmark, A. *Science* **224**, 1348-1350 (1984).
19. Baekkeskov, S. *et al. J. clin. Invest.* (in the press).
20. Harlow, E., Williamson, N. M., Ralston, R., Helfman, D. M. & Adams, T. E. *Molec. cell. Biol.* **5**, 1601-1610 (1985).
21. Tevethia, S. S., Flyer, D. C. & Tjian, R. *Virology* **107**, 13-23 (1980).
22. Pfizenmaier, K., Pan, S. & Knowles, B. B. *J. Immun.* **124**, 1888-1891 (1980).
23. Abramczuk, J., Pan, S., Maul, G. & Knowles, B. B. *J. Virol.* **49**, 540-548 (1984).
24. Sternberger, L. A. *Immunocytochemistry*, 2nd edn (Wiley, New York, 1979).

LETTERS TO NATURE

A model for collimated outflows in molecular clouds and the case of HH 7-11

G. Silvestro*, A. Ferrari*, R. Rosner†, E. Trussoni‡ & K. Tsinganos§

* Istituto di Fisica Generale dell'Università, Torino, Italy

† Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA

‡ Istituto di Cosmo-geofisica del CNR, Torino, Italy

§ Department of Physics, University of Crete, Heraklion, Greece

Collimated outflows of high-velocity gas, with typical scale lengths of 0.01-0.2 pc, are often observed to be associated with linear chains of optical emission knots¹, sometimes connected by faint nebulosities. The flows have been ascribed to the interaction between an energetic wind ejected by a young stellar object and circumstellar material. However, none of the acceleration and collimation mechanisms proposed so far can easily account for the most peculiar properties of the outflows, namely the almost constant spacing between the knots, their radial motion and the abrupt velocity changes along the flow, well illustrated by the

prototype we use for our modelling, HH 7-11. We suggest here a new interpretation of the phenomenon, based on flows in a channel of variable cross-sectional area due to Kelvin-Helmholtz instabilities between the flow and the ambient medium; and present solutions of the Mach number equation for such a channel, which possess multiple critical points and shocks identified with the observed optical knots.

Linear chains of optical emission line knots aligned with infrared sources embedded in dark clouds possess recession velocities which are interpreted as due to outflow from the infrared source. These chains often lie along the axis of high-velocity bipolar CO outflows, which have been observed in association with molecular clouds². Although such molecular flows, in general, have poor collimation (the opening angle is $<20^\circ$ in only a few sources), new data on optical jets show much lower values for opening angles ($\sim 5^\circ$ - 10°). These data also trace the flow very close (1-2 arc s) to their driving star; thus collimation must take place on distances ≤ 100 AU from the source. The wind-flow model that we propose is motivated by remarkable structural similarities between the latter well-collimated outflows in molecular clouds and jets from active galactic nuclei and quasars³, which suggest that similar hydrodynamic phenomena may occur in objects with widely different length scales (see Fig. 1). To illustrate our proposal we shall focus our attention on the chain of Herbig-Haro objects HH 7-11, which seem to be excited by a wind flowing out of the infrared source

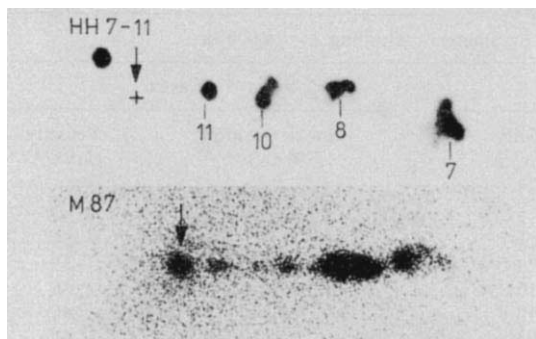


Fig. 1 The chain of optical knots HH 7-11. The arrow indicates the position of the driving infrared source. A similar morphology is observed in the optical jet of the elliptical galaxy M87.

HH 7-11 IR in the NGC1333 region. Our choice has to do with the remarkably uniform spacing of these objects, which may be a prototype for molecular outflow sources.

The NGC1333 complex is estimated to be at a distance of ~ 500 pc from Earth⁴. The molecular cloud was mapped⁵ in the (J, K) = (1, 1) line of NH_3 , and the southern condensation is estimated to have mass $\sim 230 M_\odot$. Bipolar CO gas flow associated with a low-luminosity infrared source in the densest part of the cloud has been discovered⁶ in the vicinity of HH 7-11. The peak intensities of blueshifted and redshifted wings are separated by 1.05 arc min (~ 0.20 pc), the total mass in high-velocity gas is estimated to be $4.2 M_\odot$ ($1.3 M_\odot$ in blueshifted gas), and the total kinetic energy is 7×10^{45} ergs. A map of the region in the CS(J = 3 $\rightarrow$ 2) line⁷ gives evidence for an elongated structure of high-density gas through the source of outflow, orthogonal to the separation of the 'red' and 'blue' high-velocity CO components. Measured⁴ radial velocities range from -16 to -133 km s^{-1} , with an uncertainty $\pm 15 \text{ km s}^{-1}$, for the chain of aligned objects HH 7, 8, 10 and 11, respectively, with the proper motion $\mu = 3.5 \text{ arc s per } 100 \text{ yr}$ for HH 11 (ref. 8), giving a transverse velocity $v_t = 83 \text{ km s}^{-1}$ at an angle $\psi = 154^\circ$ to the line of sight, at the source distance. The infrared source HH7-11 IR (ref. 9), is an exceptionally red object (K-L > 2.0), with an estimated lower limit to the bolometric luminosity of $\sim 60.8 L_\odot$ (ref. 10). This infrared source "has surely somehow ejected the matter exciting the chain HH 7-11"¹⁰. The parameters of the HH 7-11 chain are summarized in Table 1.

For modelling the flow over the scale length of the HH 7-11 chain, we assume the system to consist of a central condensation, plus an envelope of negligible mass. Indeed, the measured colour excess¹¹ towards HH 11 and HH 7, together with an estimate for the mean extinction in the galactic plane¹² and the extinction versus column density relation¹³, allow us to estimate that both the mass in the spherical shell between HH 11 and HH 7 and the mass in the cloud outside HH 7 are essentially negligible when compared with the total cloud mass. We shall then assume that a central mass is surrounded by a circumstellar disk in which two narrow channels are formed, and through which the collimated stellar wind flows.

The pressure of a circumstellar disk may be important in collimating the jet. For our source, detection⁷ of the CS transition J = 3 $\rightarrow$ 2 implies a density $\geq 1.6 \times 10^6 \text{ cm}^{-3}$, which suggests that the jet could be pressure confined; estimates of the density in other similar objects, such as L 1551 IRS5, give values¹⁴ $\sim 4 \times 10^6 \text{ cm}^{-3}$. Thus, we shall assume that at some base radius, say $z_0 = 10^3 \text{ AU}$, the wind is already supersonic and collimated with a uniform temperature $T_0 \approx 10^4 - 10^5 \text{ K}$, corresponding to a sound speed $c_s \approx 16 - 50 \text{ km s}^{-1}$. A flow thermally confined by the external medium is probably Kelvin-Helmholtz unstable. As studied in detail in the case of extragalactic jets¹⁵, a beam develops unstable pinching modes, whichever is its Mach number, which

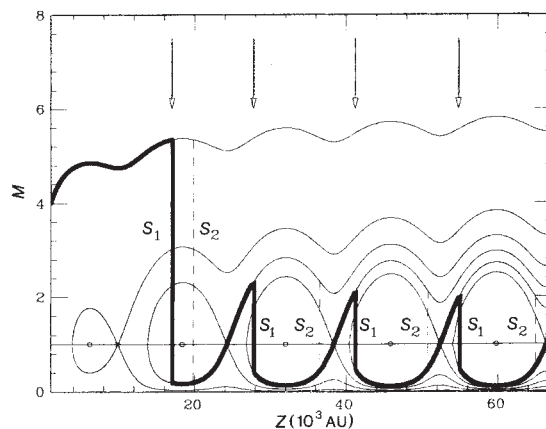


Fig. 2 Behaviour of the flow Mach number M with distance Z . The complete topology of solutions is plotted: the actual solution for the chain HH 7-11 is marked by the heavy line. The pairs of shocks S_1 and S_2 connecting different solutions are shown; the arrows indicate the positions of those shocks we associate with the objects (from the left) HH 11, HH 10, HH 8 and HH 7, respectively. The plot is given for an isothermal flow with: $m = 20 M_\odot$, $T_0 = 2 \times 10^4 \text{ K}$, $\epsilon_0 = 0.45$, $M_0 = 4$ and $\lambda = 14,000 \text{ AU}$.

lead to compressions and rarefactions of the gas periodically distributed (depending on the perturbation wavelength λ) along the jet. Subsequent propagation along the channel will then be governed by the equations of mass and momentum conservation, that in a quasi-two-dimensional approximation¹⁶ can be combined into a single equation for the Mach number $M = v/c_s$ (with $Z = z/z_0$),

$$\frac{M^2 - 1}{2M^2} \frac{dM^2}{dZ} = \frac{d \ln A(Z)}{dZ} - \frac{Gm}{2c_s^2 Z^2 z_0} \quad (1)$$

where $A(Z)$ is the variable cross-sectional area of the channel, $A(Z) = Z^2 [1 + \epsilon_0 \sin(2\pi Z z_0 / \lambda)]^2$. In deriving these expressions, we have adopted spherical symmetry, that is the beam is assumed to be conical with fixed opening angle, modulated by a pinching perturbation with wavelength λ , and amplitude ϵ_0 at $Z = 1$. The quasi-two-dimensional approximation used here is only valid for causally connected streamlines, that is, for the flow along an infinitesimal flow tube lying on the axis of the beam. As is usual in such quasi-two-dimensional treatments^{17,18}, we adopt the rate of geometric variation of the macroscopic flow channel as an appropriate measure of the cross-sectional areal variation of the elemental flow tube along the channel axis. In the present case, these areal variations imply periodic subtraction (for contraction) or input (for expansion) of momentum to the flow, which can lead to the formation of regularly-spaced standing shocks. As heating of gas is expected in a shock, we propose that a sequence of such shocks can be associated with the observed HH chain. The limits of this interpretation depend on our 'slender tube' approximation; hence, we cannot account for the possible effects of oblique shocks (which are connected with the interaction of the beam surface with the external medium). Thus, strictly speaking, we cannot use our very simple model

Table 1 Distances from HH 7-11 IR and velocities of the knots of the chain HH 7-11

	θ (")	Distance (AU)	Velocity (km s^{-1})
HH 11	12.4	12,500	168
HH 10	26.6	26,500	33
HH 8	40.2	40,500	50
HH 7	62.1	63,000	67

Table 2 Positions of shocks and velocities for the following parameters (assuming $\lambda = 14,000$ AU)

Case a		Case b		Case c	
Shock position (AU)	Velocity (km s ⁻¹)	Shock position (AU)	Velocity (km s ⁻¹)	Shock position (AU)	Velocity (km s ⁻¹)
16,990	126	16,370	317	17,250	233
19,810	126	20,210	317	20,120	233
27,750	55	27,430	120	28,730	55
36,490	55	36,730	120	35,930	55
41,170	51	40,480	110	42,010	46
50,750	51	51,230	110	50,290	46
54,910	47	54,250	104	55,610	41
64,170	47	65,350	104	64,490	41

Case a: $m = 20 M_{\odot}$, $T_0 = 2 \times 10^4$ K, $\varepsilon_0 = 0.45$, $M_0 = 4$, $\gamma = 1$.

Case b: $m = 100 M_{\odot}$, $T_0 = 1 \times 10^5$ K, $\varepsilon_0 = 0.5$, $M_0 = 5$, $\gamma = 1$.

Case c: $m = 20 M_{\odot}$, $T_0 = 1 \times 10^5$ K, $\varepsilon_0 = 0.3$, $M_0 = 4$, $\gamma = 4/3$.

to directly infer phenomena on resolved transverse dimensions, for which complete two-dimensional numerical simulations are required.

In principle, we ought to take into account the velocity of the perturbations along the surface of the jet; however, for the fastest growing perturbations this velocity is quite small so that the areal distortions can be considered localized without loss of generality¹⁹. Solutions of equation (1) have been found for the case of HH 7-11, assuming $\lambda \approx 14,000$ AU, the initial $M(Z = 1) = M_0 > 1$, and $\varepsilon_0 (< 1$ always) to be a free parameter of the solutions. Figure 2 shows a plot of a representative solution for a central mass $m = 20 M_{\odot}$, $T_0 = 2 \times 10^4$ K, $\varepsilon_0 = 0.45$. The main feature of interest is that with $A(Z)$ given as above, we find multiple critical points and shocked solutions. The position of each shock is determined by the Rankine-Hugoniot condition for conservation of mass and momentum fluxes across the discontinuity for an isothermal gas, $M_1 M_2 = 1$, where M_1 and M_2 are the Mach numbers upstream and downstream from the shock, respectively: every critical solution can be connected with the following one by two shock transitions (with the same velocity) at the positions S_1 and S_2 . In Table 2 (case a), we report the position of the shocks with the corresponding upstream flow velocity: we see that the first of each pair of shocks (S_1) can be tentatively associated with the observed morphology; in particular, we associate the first fast moving shock with HH 11, and the slower shocks with HH 10, HH 8 and HH 7, respectively. The essential question is then whether these shocks actually form; the answer to this crucial question can only be found by a time-dependent treatment of the problem because the present stationary solutions select only the positions where the shocks can form.

The dependence of the results on the value of the central mass and on the assumption of uniform temperature flow was explored by modelling the cases of an isothermal flow with $m = 100 M_{\odot}$, $T_0 = 10^5$ K, $\varepsilon_0 = 0.60$ (case b) and a polytropic flow ($\gamma = 4/3$) with $m = 20 M_{\odot}$, $T_0 = 10^5$ K and $\varepsilon_0 = 0.30$ (case c). The position of the shocks is almost the same in the three cases and again fits the position of the HH objects (Table 2). However, a larger central mass requires a higher temperature to accelerate the flow, and also a larger beam areal disturbance amplitude to produce the shocks (case b): therefore a central mass $\gg 20 M_{\odot}$ implies velocities larger than those observed. A slight departure from isothermal conditions leads to velocities similar to the isothermal case, but for lower amplitudes of perturbations, that is, shock formation is obtained under less restrictive conditions in polytropic flows (case c). Furthermore, if a diffuse mass comparable or larger than the central one is present in the outflow region, our results must be recalculated because the flow can be severely affected by a spatially distributed mass: for specific radial mass distribution laws, no stationary outflow

solutions may exist¹⁹.

Our interpretation of the HH 7-11 chain as arising from the interaction of a collimated outflow with dense circumstellar material agrees with well-established properties of the HH nebulae, whose ionization and excitation appear to be dominated by collisional processes in shock waves with shock velocities ~ 100 km s⁻¹. Furthermore, proper motion data of HH objects, and their alignment with the infrared source in several cases, suggest that the supersonic flows responsible for excitation are highly collimated²⁰. New observational data²¹ show several bright HH objects to be associated with a continuous jet. For the HH 7-11 chain, a diffuse spatially extended emission in the S_1 line of H_2 was detected²², and interpreted as arising from the same low-velocity shock that causes visible emission. Similar results²³ were obtained for the same regions in the $v = 1 \rightarrow 0$ Q-branch line of H_2 . In addition, weak $H\alpha$ emission has been recently detected²⁴ between the knots of the HH 7-11 chain. These data are consistent with a model in which HH knots are powered by highly collimated quasi-continuous flows, the knots themselves marking the location of increased energy dissipation in the shock fronts occurring along the flows. The observation²⁵ of abrupt velocity changes over very small spatial elements in the bipolar gas flow surrounding L1551IRS5 supports our model.

We thank Dr R. Mundt for material in advance of publication. This work was supported by the Italian Ministero della Pubblica Istruzione and the Consiglio Nazionale delle Ricerche, and by the NSF grant AST83-03522. G.S., A.F. and E.T. thank the Center for Astrophysics for hospitality.

Received 11 April; accepted 14 November 1986.

- Mundt, R. & Fried, J. W. *Astrophys. J.* **274**, L83-L86 (1983).
- Bally, J. & Lada, C. J. *Astrophys. J.* **265**, 823-847 (1983).
- Ferrari, A., Habbal, S. R., Rosner, R. & Tsinganos, K. *Astrophys. J.* **277**, L35-L39 (1984).
- Strom, S. E., Grasdalen, G. L. & Strom, K. M. *Astrophys. J.* **191**, 111-142 (1974).
- Ho, P. T. P. & Barrett, A. H. *Astrophys. J.* **237**, 38-54 (1980).
- Snell, R. L. & Edwards, S. *Astrophys. J.* **251**, 103-107 (1981).
- Schwartz, P. R., Waak, J. A. & Smith, H. Z. *Astrophys. J.* **267**, L109-L114 (1983).
- Herbig, G. H. & Jones, B. F. *Astr. J.* **88**, 1040-1052 (1983).
- Strom, S. E., Vrba, F. J. & Strom, K. M. *Astr. J.* **81**, 314-316 (1976).
- Cohen, M. & Schwartz, R. D. *Astrophys. J.* **265**, 877-900 (1983).
- Bohm, K. H., Brugel, E. W. & Holmsted, E. *Astr. Astrophys.* **125**, 23-28 (1983).
- Spitzer, L. Jr *Diffuse Matter in Space* (Interscience, New York, 1968).
- Bohlin, R. C., Savage, B. D. & Drake, J. F. *Astrophys. J.* **224**, 132-142 (1978).
- Snell, R. L., Bally, J., Strom, S. E. & Strom, K. M. *Astrophys. J.* **290**, 587-595 (1985).
- Ferrari, A., Trussoni, E. & Zaninetti, L. *Mon. Not. R. Astr. Soc.* **196**, 1051-1066 (1981).
- Parker, E. N. *Interplanetary Dynamical Processes* (Interscience, New York, 1963).
- Kopp, R. A. & Holzer, T. E. *Sol. Phys.* **49**, 43-56 (1976).
- Holzer, T. E. *J. geophys. Res.* **82**, 23-35 (1977).
- Ferrari, A., Trussoni, E., Rosner, R. & Tsinganos, K. *Astrophys. J.* **300**, 577-590 (1986).
- Schwartz, R. D. *Rev. Mex. Astr. Astrofis.* **7**, 27-54 (1983).
- Mundt, R., Buhrke, T., Fried, J. W., Neckel, T., Sarcander, M. & Stocke, J. *Astr. Astrophys.* **140**, 17-23 (1984).
- Zealey, W. J., Williams, P. M. & Sandell, G. *Astr. Astrophys.* **140**, L31-L34 (1984).
- Lightfoot, J. F. & Glencross, W. M. *Mon. Not. R. Astr. Soc.* **221**, 993-1000 (1986).
- Mundt, R. *Can. J. Phys.* **64**, 407-413 (1986).
- Fridlund, C. V. M., Sandqvist, A., Nordh, H. L. & Olofsson, G. *Astr. Astrophys.* **137**, L17-L19 (1984).

Cometary magnitude distribution and the fading of comets

David W. Hughes

Department of Physics, The University, Sheffield S3 7RH, UK

Long-period comets are, on average, more luminous than short-period ones. This paper considers the recent observations of the nucleus of comet Halley, in conjunction with the results of the modelling of the accretion process that is most probably responsible for cometary origin. It now seems that simple physical processes such as a rounding off and removal and/or settling of bumps and prominences protruding from the nucleus plus a change in the percentage of the nucleus surface that is active could account for the fading that occurs just after jovian cometary capture.

A series of observations of the apparent magnitude, m , of a specific comet fit well a formula of the type

$$m = H_{10} + 2.5 n \log r + 5 \log \Delta \quad (1)$$

where r and Δ are the comet-Sun and comet-Earth distances and H_{10} , the absolute magnitude, is a product of the fitting in the specific case where n is forced to be 4.0. A file of H_{10} values has been produced using data published by Vsekhsvyatskii¹⁻⁵ and Vsekhsvyatskii and Il'ichishina⁶. The accuracy of specific H_{10} values is enhanced for long-period comets by the fact that 60% of this group have perihelion distances, q , < 1 AU and so H_{10} can be obtained by interpolation using equation (1). Of known long-period comets, 89% have $q < 2.0$ AU (ref. 7). Only 19% of short-period comets have $q < 1.0$ AU (85% have $q < 2.0$ AU) but this is partially compensated for by the fact that most of these have been seen more than once, leading to an increase in the accuracy of H_{10} . A plot of the logarithm of the number of comets, N , observed up to the present to be brighter than a specific absolute magnitude, H_{10} , against that magnitude (Fig. 1) produces a cumulative curve where

$$N = ba^{H_{10}} \quad (2)$$

to the right of a knee value. To the left of that value the data

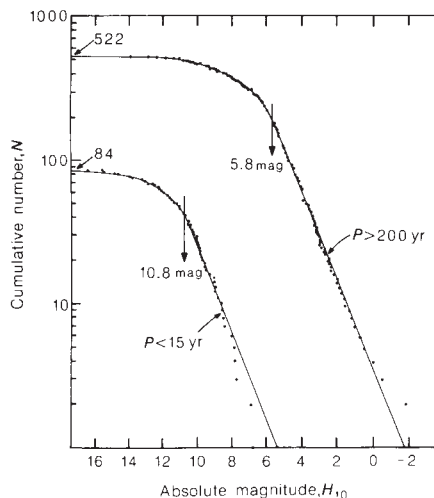


Fig. 1 The cumulative number, on a logarithmic scale, of short period ($P < 15$ yr) and long period ($P > 200$ yr) comets as a function of their absolute magnitude, H_{10} (see equation (1)). The line through the $P > 200$ yr data has been obtained by linear regression fitting and the line through the $P < 15$ yr has been purposely drawn such that its gradient is the same as that of the long period line. The fit to these data is good. The data break away from linearity at 5.8 mag and 10.8 mag respectively. Mankind has so far recorded 522 long-period and 84 short-period comets sufficiently well to obtain good absolute magnitude values.

are greatly affected by observational selection because of Earth-based observers missing faint comets.

These data have been divided crudely into two groups, long period comets (with period, $P > 200$ yr) which have been seen only once and have probably been recently captured from the Oort/Öpik cloud and are evolutionarily 'new', and short period comets (period, $P < 15$ yr) which have been captured into the inner solar system by Jupiter, have a low orbital inclination, a median period of ~ 7 yr and are evolutionarily 'old'—being smaller and less massive than they originally were.

Four things are immediately noteworthy in Fig. 1. The right hand, linear portions of both curves essentially have the same gradient, being represented by

$$\log N = 0.549 + 0.301 H_{10} \quad (P > 200 \text{ yr}) \quad (3)$$

$$\log N = -1.625 + 0.301 H_{10} \quad (P < 15 \text{ yr}) \quad (4)$$

The knee in the long period curve occurs at 5.8 mag whereas the short period knee is at 10.8 mag. The vertical separation of the two right hand portions is equivalent to a factor of 150. This contrasts with the fact that in recorded scientific history we have noted 522 long period comets and 84 short period comets, a ratio of 6.2:1.

How do we explain these observations? Let us first move a step beyond the absolute magnitude. For a well-behaved periodic middle-aged comet such as Halley it is reasonable to assume that the cometary luminosity is proportional to the surface area of the nucleus^{8,9} and thus produce^{10,11} formulae of the type

$$\log R = 1.842 - 0.2 H_{10} \quad (5)$$

and

$$\log M = 20.845 - 0.6 H_{10} \quad (6)$$

where R (km) is the equivalent radius of the nucleus (the radius of a sphere with the same surface area) and M (g) is the mass of the nucleus. The only cometary nucleus that has been imaged to date is that of Halley. It is relatively smooth and potato shaped, and on March 14 1986 was emitting gas and dust profusely from about 8% of its surface area¹². (The adjective 'well-behaved' is applied to Halley because observations of the last thirty apparitions have indicated that its H_{10} value at each return¹³ has remained reasonably constant at $\sim 5.5 \pm 0.7$.) If it is assumed that a typical comet loses a layer of constant thickness from its active regions each time it passes the Sun and that the ratio between the active and inactive area remains constant, the decay can be easily modelled¹⁰. The process is slow and as the comet ages through 10, 20, 30, 40 and 50% of its inner solar system life its mass drops to 73, 51, 34, 22 and 13 percent respectively of its initial value and its absolute magnitude, which on capture was H , becomes $H + 0.23$, $H + 0.48$, $H + 0.77$, $H + 1.1$ and $H + 1.5$ respectively. It can thus be concluded that a capture and decay, which keeps constant both the general shape and the percentage of the surface area that is active, changes the absolute magnitude by only a small amount. The presently accepted mechanism¹⁴⁻¹⁶ for transferring a long period comet into a short period orbit, a mechanism which requires only a few near tangential close passages of the jovian orbit, will, if the above assumptions are correct, do little to change the absolute magnitude of a comet. So there is no *a priori* reason why a specific group of long period comets should have a median absolute magnitude that differs significantly from the group of short period comets that they are perturbed into.

Where have we gone wrong? Why do comets fade between their long and short period existence? Previous suggestions relied on chemistry, having new long period comets coated with some supervolatiles. After a few close passages of the Sun these had been lost leaving behind the more mundane mixture of H_2O , CO_2 , NH_3 , CH_4 and so on.

But why not simply rely on physics? The magnitude distribution index, a , in equation 1, is 2.0 ± 0.2 (see equations (3) and (4)). Using equation (6) this can be converted to a mass distribution index, s , of 1.75 ± 0.1 . (This index is defined such that the

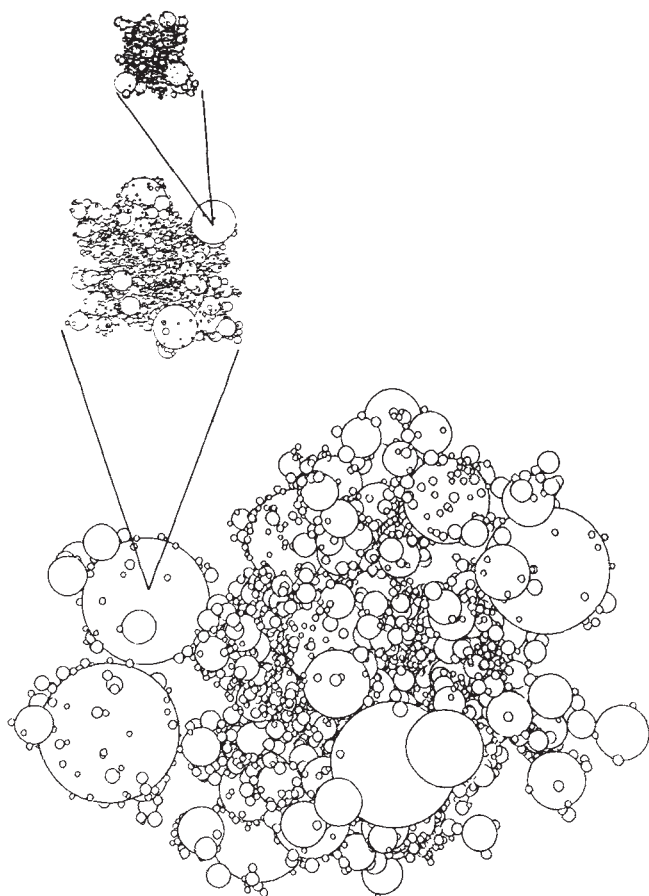


Fig. 2 A fractal model of a newly accreted cometary nucleus. This is a hierarchical porous, low density structure which is, as shown, self similar over a vast range of scale lengths from kilometric down to micrometric.

number of comets with mass greater than M is proportional to $M^{(1-s)}$. Such a low mass distribution index is indicative¹⁷ of the fact that accretion, as opposed to fragmentation, is the main agent responsible for formation. The resultant comet will have a fractal form^{18,19} such that its hierarchical, porous structure will be nearly self similar over a vast range of scale lengths from kilometric down to micrometric. The fractal model of a newly accreted comet is illustrated in Fig. 2. The hummocky nature of this model differs from the rather smooth shape presented by the images of Halley. The first few solar close approaches of a new comet could result in a general rounding off of its features and a sharp decrease in the surface area to mass ratio. As luminosity has been assumed to be proportional to surface area this could result in a considerable change in the absolute magnitude. A sharp drop in surface area by a factor of say 2.5 changes the magnitude by 1.

Only 8% of the surface of Halley (a slowly rotating nucleus with a spin period of about 52 h) is active at any one time. Maybe a new comet has not had time to build up choking surface dust regions and thus all the sunward face of the nucleus, about 50% of its total area, is actively emitting gas and dust. The ratio 50/8 accounts for another 2 magnitudes.

Both these proposed effects change the magnitude by a total of say ΔH and equation (6), for new as opposed to middle-aged comets, becomes $\log M = (20.845 - 0.6 \Delta H) - 0.6 H_{10}$. This is equivalent to sliding the $P > 200$ yr curve in Fig. 1 to the left by ΔH and thus considerably decreasing the difference between the observed numbers of long and short period comets. The vertical separation between the two right hand portions of the cumulative curves (see Fig. 1) would then be $10^{(2.174 - 0.301 \Delta H)}$. So a ΔH of 3 would change this factor from 150 to 19.

Returning to the other three noteworthy points in Fig. 1, the

similarity in the gradients of the two curves is not at all surprising²⁰ if it is assumed that decay is equivalent to the loss of a constant thickness of a constant percentage of the nucleus surface at each apparition as opposed to a constant percentage of the mass. The observational bias inherent in cometary observation does not markedly affect the gradients of these two curves. Dividing the $P < 200$ yr comets into specific period groups leads to a set of a values scattered randomly about the mean value given above. Dividing the $P > 200$ yr comets into groups with differing perihelion distances again leads to a set of a values randomly scattered about the mean (see Hughes²¹).

The difference between the two knee positions, 10.8 mag for short-period comets and 5.8 mag for long-period comets, is equivalent to a factor of 100 in brightness. The discussion above enables us to change this 5 mag difference to $(5 - \Delta H)$ and a ΔH of 3 gives a factor of 6 in brightness. This can easily be accounted for by the differences in the probability of detection²² between the two groups. The fact that we see a long-period comet only once and a short-period comet returns, usually along the ecliptic, on average every 7 yr underlines this difference.

We are still left with the problem of calculating the actual ratio between the number of long-period and short-period comets that enter a specific sun-centred volume of the solar system and have masses greater than a specific mass. Both the differences in observational probability and the differences in the absolute magnitude-mass relationships have to be taken into account. From Fig. 1 we can conclude that this ratio is ≤ 150 but is more likely to be in the region of 20. The fact that the apparent value of this ratio is 6.2 means that many long-period comets fly past Earth undetected.

Received 8 September 1986; accepted 13 November 1986.

1. Vsekhsvyatskii, S. K. in *Fizicheskie Kharakteristiki Komet* (Moscow, 1968) (Translated as *Physical Characteristics of Comets*, Israel Program for Scientific Translations, Jerusalem, 1964).
2. Vsekhsvyatskii, S. K. *Soviet Astr.* **6**, 849-854 (1963).
3. Vsekhsvyatskii, S. K. *Soviet Astr.* **8**, 429-431 (1964).
4. Vsekhsvyatskii, S. K. *Soviet Astr.* **10**, 1034-1041 (1967).
5. Vsekhsvyatskii, S. K. in *The Motion, Evolution of Orbits and Origin of Comets* (ed. Chebotarev, G. A. & Kazimirchak-Polonskaya, E. I.) 9-15 (Reidel, Dordrecht, 1972).
6. Vsekhsvyatskii, S. K. & Il'ichishina, N. I. *Soviet Astr.* **15**, 310-313 (1971).
7. Hughes, D. W. *Contemp. Phys.* **23**, 257-283 (1982).
8. Whipple, F. L. *Mem. Soc. R. Sci. Liege Ser. 6* **9**, 101-111 (1976).
9. Öpik, E. J. *Astrophys. space Sci.* **21**, 307-398 (1973).
10. Hughes, D. W. *Mon. Not. R. astr. Soc.* **213**, 103-109 (1985).
11. Hughes, D. W. *Phil. Trans. R. Soc.* (in the press).
12. Keller, H. U. et al. *Nature* **321**, 320-326 (1986).
13. Hughes, D. W. *Mon. Not. R. astr. Soc.* **204**, 1291-1295 (1983).
14. Everhart, E. *Astrophys. Lett.* **10**, 131 (1972).
15. Fernandez, J. A. *Astr. Astrophys.* **96**, 26 (1981).
16. Fernandez, J. A. & Ip, W. A. in *Asteroids, Comets and Meteors* (ed. Lagerkvist, C. I. & Rickman, H.) 387-390 (University of Uppsala Press, 1983).
17. Daniels, P. A. & Hughes, D. W. *Mon. Not. R. astr. Soc.* **195**, 205-212 (1981).
18. Donn, B. & Hughes, D. W. *Bull. Am. astr. Soc.* **17**, 689 (1985).
19. Daniels, P. A. & Hughes, D. W. *Mon. Not. R. astr. Soc.* **195**, 1001-1009 (1981).
20. Hughes, D. W. & Daniels, P. A. *Mon. Not. R. astr. Soc.* **198**, 573-582 (1982).
21. Hughes, D. W. *Mon. Not. R. astr. Soc.* (in the press).
22. Kresák, L. in *Comets* (ed. Wilkening, L. L.) 56-82 (University of Arizona Press, 1982).

Study of particle motion in concentrated dispersions by tracer diffusion

R. H. Ottewill & N. St. J. Williams

School of Chemistry, University of Bristol,
Bristol BS8 1TS, UK

Monodisperse spherical colloidal particles can be used to simulate molecular motions and, in particular, those which occur in the regions of phase transitions such as melting and freezing. In nonaqueous media the behaviour of sterically stabilized particles in near to that of 'hard-spheres'. The use of systems in which the majority of the particles are refractive index matched to the dispersion medium but where a few particles are added of the same

size and surface chemistry but of different refractive index allows the development of tracer particle measurements using photon correlation spectroscopy. Thus it becomes possible to examine particle motion in dense systems. Here we use this novel technique to determine particle self-diffusion coefficients over a wide range of volume fractions and into the 'frozen' state.

Studies on the physics of the liquid state and the phase transitions between liquids and solids have been hampered by the difficulties of labelling molecules in order to study their diffusional motion. However, one of the major developments in the field of colloid science over the past few decades has been the emergence of synthetic methods for the preparation of spherical polymer particles with a very narrow size distribution, that is, polymer latices or polymer colloids¹. By the use of these particulate systems it has become possible to study the behaviour of dispersions of high volume fraction by small-angle neutron scattering² and light diffraction³. These systems, which are of interest in their own right for the study of particle-particle interactions in colloidal dispersions, also simulate the 'liquid state' and at volume fractions in the region of 0.5 enable the liquid-solid phase transitions for 'hard spheres' to be observed^{4,5}.

Although light forms an excellent probe for the examination of dilute strongly interacting systems⁶ the interpretation of results from concentrated aqueous dispersions is complicated by multiple scattering because of the large difference between the refractive index of the particle and the medium. With aqueous based polymer latices little opportunity exists to vary the refractive index of the dispersion medium without changing the colloidal properties of the system. The development of methods for forming polymer latices in nonaqueous media⁷, however, introduces wide possibilities in the choice of the dispersion medium. In earlier studies refractive index matching between particles and the medium was achieved using a mixture of dodecane and carbon disulphide^{8,9}; matching was achieved over a narrow 'window' of wavelength and temperature although some swelling occurred on addition of carbon disulphide¹⁰. The particles used were composed of poly(methylmethacrylate) (PMMA) cores with chains of poly(12-hydroxystearic acid) chemically grafted to the core. Particles of this type are said to be sterically stabilized, that is, they interact with a short range repulsive potential essentially as 'hard spheres'^{5,7}.

The possibility of refractive index matching leads immediately to the prospect of developing 'tracer' methods for the examination of particle motion in dispersions using photon correlation spectroscopy and its feasibility has been demonstrated by Kops-Werkhoven *et al.*^{11,12}, using dispersions of silica particles in cyclohexane. These authors exploited the heterogeneity of the silica particles and hence their variation in refractive index as a function of temperature to obtain a few results for particulate diffusion at high volume fractions.

The technique is, however, capable of considerable versatility by a suitable choice of particles and medium. Moreover, by the correct choice of particle size both 'short-time' and 'long-time' diffusion coefficients can be determined. In this communication we report the use of the photon correlation technique to study the diffusion of tracer particles of PMMA in a wide range of volume fractions of polyvinyl acetate (PVA); both particles were stabilized by poly(12-hydroxystearic acid). The PVA particles had a refractive index of 1.471 and could be precisely optically matched using a mixture of *cis*-decalin (refractive index 1.481) and *trans*-decalin (1.469). PMMA particles of refractive index 1.483 were added in a ratio of 1:43 PVA particles. The hydrodynamic radius of the PMMA particles was 83 ± 2 nm and that of the PVA particles, 85 ± 2 nm. It was therefore assumed, as the same stabilizer was used and the particle sizes were very close, that the repulsive interactions between the particles were identical.

With this system it has proved possible to carry out tracer diffusion studies over a wide range of scattering vectors (K),

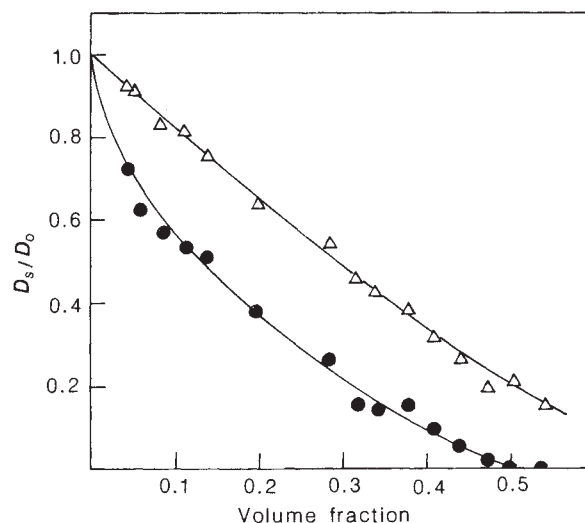


Fig. 1 Ratio of self-diffusion coefficient, D_s , to the single particle diffusion coefficient D_0 , as a function of volume fraction: Δ , short-time self-diffusion; $\bullet$, long-time self-diffusion.

correlation delay times (τ) and volume fractions. The volume fractions examined ranged from dilute, where essentially free Brownian motion occurred, through the intermediate 'liquid-like' region, up to 0.54 just above the suggested onset of freezing for a hard-sphere system. As the volume fraction increases and particle-particle interactive contacts increase the diffusional motion of the particles can be separated into self-diffusion and collective diffusion¹³. The tracer technique measures the self-diffusion of particles and further, by making measurements at short correlation times, such that the particle has only moved a distance corresponding to a small fraction of a particle radius, the short-time diffusion coefficient can be obtained. On a time scale such that the particle diffuses over a distance equivalent to several diameters and interacts both colloiddally and hydrodynamically with other particles then a long-time diffusion coefficient can be determined¹³. For angular light scattering with a scattering vector, $K = 10^{-2} \text{ nm}^{-1}$, the spatial dimension probed ($2\pi/K$) is approximately 610 nm in comparison with a particle radius of 85 nm; thus, the time for a particle to diffuse one radius is ~ 1 ms. As suggested by Pusey and Tough¹⁴ by making measurements at small scattering vectors and long correlation times it is possible to measure long-time self-diffusion coefficients in a particle ensemble. The tracer technique described allows this to be done and in the present work delay times up to 70 ms were utilized.

The results of the normalized long-time self-diffusion coefficients, D_s/D_0 , of the tracer PMMA particles are shown in Fig. 1 as a function of the volume fraction of the dispersion. D_s/D_0 decreases with volume fraction and tends to a very small value as the volume fraction tends towards 0.54. It is of interest to note that Hoover and Ree¹⁵ predicted some time ago that a freezing transition for a hard-sphere system should occur at a volume fraction of 0.495 and melting at 0.545. More recently, Pusey and van Megen⁵ reported from visual observations on refractive index matched PMMA dispersions that melting occurs at an effective hard-sphere volume fraction of 0.536. From Fig. 1 it is apparent that the onset of virtually zero diffusional motion is very close to the predicted freezing value of 0.495. It is possible that some motion still appears above this volume fraction but if it does, it is on a very long time-scale. Effectively, the long-term self-diffusional motion, which is movement through the lattice, appears to cease at the freezing transition. The short-time diffusional coefficient, however, which represents the 'rattling' motion of a particle on a lattice site remains finite in this region⁹ and as suggested by Pusey¹³ is larger than the long-time value.

This is clearly illustrated in Fig. 1 where the results for short-time self-diffusion are also plotted. The theoretical relationship between short-time self-diffusion has been developed theoretically and the results shown in Fig. 1 give good agreement with the multibody treatment of Beenakker and Mazur¹⁶, whereas that for the long-time self-diffusion behaviour is not known beyond the first term in concentration¹³ which predicts a less strong dependence on volume fraction. It is to be hoped that this detailed set of experimental results will stimulate further theoretical work in this area.

The results obtained demonstrate the use of monodisperse nearly 'hard-sphere' colloidal particles to simulate physical processes. The tracer technique used in conjunction with photon correlation spectroscopy offers considerable promise as a means of examining dynamic events occurring in concentrated colloidal dispersions.

We thank Dr P. N. Pusey for a number of valuable discussions and SERC for support of this work.

Received 22 August; accepted 2 December 1986.

1. Piirma, I. *Emulsion Polymerization* (Academic, New York, 1982).
2. Cebula, D. J. *et al. Faraday Discuss. Chem. Soc.* **76**, 37-52 (1983).
3. Goodwin, J. W., Ottewill, R. H. & Parentich, A. *J. phys. Chem.* **84**, 1580-1586 (1980).
4. Hachisu, S. & Takano, K. *Adv. Colloid Interface Sci.* **16**, 233-252 (1982).
5. Pusey, P. N. & van Megen, W. *Nature* **320**, 340-342 (1986).
6. Brown, J. C., Pusey, P. N., Goodwin, J. W. & Ottewill, R. H. *J. Phys. A* **8**, 664-682 (1975).
7. Cairns, R. J. R., Ottewill, R. H., Osmond, D. W. J. & Wagstaff, I. *J. Colloid Interface Sci.* **54**, 45-51 (1976).
8. Cebula, D. J., Ottewill, R. H. & Pusey, P. N. *SERC Rep.* 1982.
9. Pusey, P. N. & van Megen, W. *J. Phys. Paris* **44**, 285-291 (1983).
10. Livsey, I. & Ottewill, R. H. *Polymer* (in the press).
11. Kops-Werkhoven, M. M. & Fijnaut, H. M. *J. chem. Phys.* **77**, 2242-2253 (1982).
12. Kops-Werkhoven, M. M., Pathmanathan, C., Vrij, A. & Fijnaut, H. M. *J. chem. Phys.* **77**, 5913-5922 (1982).
13. Pusey, P. N. in *Physics of Amphiphiles: Micelles, Vesicles and Microemulsions* (eds DeGiorgio, V. & Corti, M.) 152-159 (North-Holland, Amsterdam, 1985).
14. Pusey, P. N. & Tough, R. J. A. *Adv. Colloid Interface Sci.* **16**, 143-159 (1982).
15. Hoover, W. G. & Ree, F. H. *J. chem. Phys.* **49**, 3609-3617 (1968).
16. Beenakker, C. W. J. & Mazur, P. *Physica* **120A**, 388-410 (1983); **126A**, 349-370 (1984).

Phase changes during elongational flow of polymer solutions

J. Ferguson, N. E. Hudson, B. C. H. Warren* & A. Tomatarián

Fibre and Textile Research Unit, Department of Pure and Applied Chemistry, University of Strathclyde, Glasgow G1 1XL, UK

Protein solutions with a high content of keratin have a tendency to gel on standing. Recent results with synthetic polymers have shown that a shear field can generate the same phenomenon, even in solutions that do not gel at rest. We report here that in a strong elongation flow field, a gel can be induced in polymer solutions that do not gel at rest, or even in a shear field. This throws light on the process of fibre formation by the wet, dry or melt processes, and in addition, gives a possible explanation of the ability of the spider to hang from a freshly extruded thread.

Trouton¹ showed 80 years ago that for newtonian fluids, the elongational viscosity was three times the shear viscosity. However, for non-newtonian fluids, this relationship only applies at the limit of very low deformations. Many polymer solutions show the characteristics of shear thinning (Fig. 1). The elongational behaviour of such solutions is more complicated. Typically, the elongational viscosity will increase, pass through a maximum, and then decrease as strain (or rate of strain) increases. When a polymeric fluid has spinnability², the extensional flow may well be characterized by a further increase in elongational viscosity, as the strain asymptotically approaches

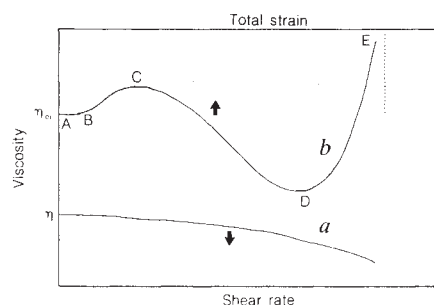


Fig. 1 Typical viscosity curves for a non-newtonian polymer fluid. Curve *a*, shear thinning behaviour, the shear viscosity falling as shear rate increases. Curve *b*, the response in an elongational flow field, divided into distinct regions. AB, Newtonian deformation, $\eta_{el} = 3\eta$; BC, 'strain hardening' deformation of the network structure within the fluid; CD, viscoelastic deformation, with short to medium lifetime entanglement junctions decreasing, causing the fall in η_{el} ; DE, elastic deformation of the network held by very long lifetime entanglements (in this region, the fluid is showing solid-like elastic deformation up to a limiting value of strain where cohesive fracture will occur).

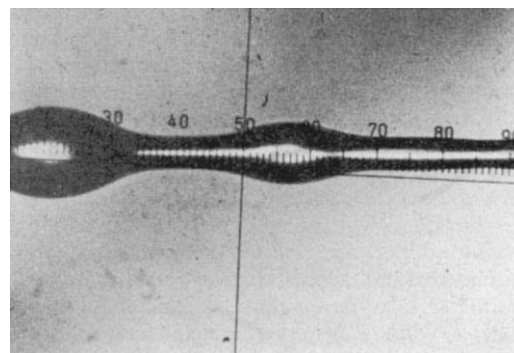


Fig. 2 Filament and droplets of 2% FM9 Copolymer in DPM after elongation had ceased. Viewed under polarized light in a wild M21 microscope system. Each division represents 40 μm .

a limiting value. This allows the fluid to fulfil the conditions for spinnability set out by Lodge², that the fluid should get stronger faster than it is getting thinner:

$$\frac{\partial \eta_{el}}{\partial \epsilon} > \frac{\partial A}{\partial t} \quad (1)$$

where η_{el} is the elongational viscosity, ϵ the total (Hencky) strain imposed, A the cross-sectional area, at time t .

It had been assumed that this rise in elongational viscosity was due entirely to long lifetime entanglements and increased hydrodynamic volume caused by the uncoiling of macromolecules in the extensional flow field. Recent results³ show that an additional mechanism can have a part to play and may dominate extensional flow behaviour in certain polymeric systems.

In an experiment rather similar to that described by Trouton, a drop of polymer solution (2% FM9 copolymer in dipropylene glycol monoethyl ether, DPM) was allowed to fall from a burette. The drop fell under gravity, dragging a filament of solution behind it. This filament remained attached to the lip of the burette, and consequently was elongated. The flow was suddenly stopped by the drop hitting a flat surface. The maximum strain imposed was 2.7 units. It was observed that droplets then formed on the filament. These droplets were stable during the timescale of the experiments (up to 6 min). In certain cases, coalescence occurred between two or more droplets. The filament, with attendant droplets, was caught on a modified microscope slide, which held the filament without allowing it to come into contact

* Present address: Chemical Defence Establishment, Porton Down, Salisbury, Wilts SP4 0JQ, UK.

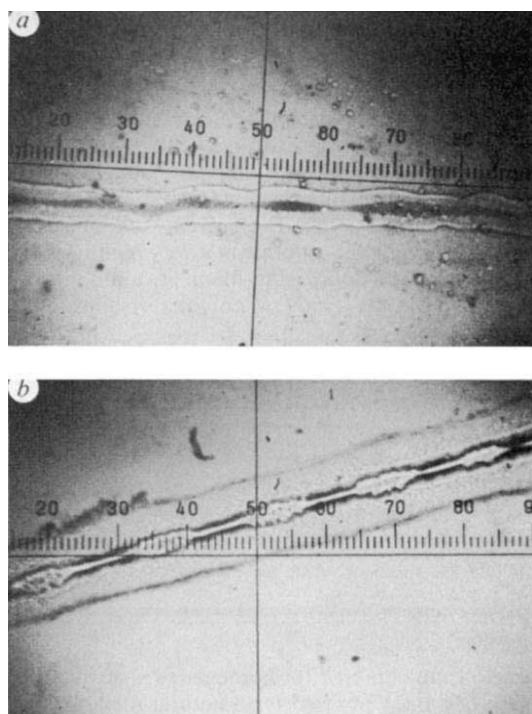


Fig. 3 *a*, Elongated filament of 2% FM9 Copolymer in DPM. Viewed under non-polarized light immediately after being placed on the microscope slide. Each division represents 40 μm . *b*, As *a*, viewed under polarized light, 24 h after being placed on the microscope slide. Each division represents 25 μm .

with the glass surface. The sample was then photographed using a Wild M21 polarizing microscope system. Figure 2 shows a line of high birefringence passing through the centre of an apparently isotropic fluid. The incipient droplets can be seen. In fluid filaments which did not form droplets when elongation ceased, no such line of birefringence could be seen.

In a separate series of experiments, the filament was rapidly extended directly onto the microscope stage, held in extension, and then photographed under non-polarized light. Although there is a slight distortion of the filament, nevertheless, as Fig. 3*a* shows, even with non-polarized transmitted light, a core of gelled polymer could be clearly observed. This indicates, of course, significantly different refractive indices of the gel and of the surrounding fluid. The solvent, DPM, had a boiling point of 188.3 $^{\circ}\text{C}$ at 760-mm pressure, and a partial vapour pressure of 0.36 mm Hg at 25 $^{\circ}\text{C}$, and so did not evaporate to any great extent. After 24 h, it was observed, Fig. 3*b*, that redissolution of the highly orientated filament was occurring. The solvent-rich phase, being largely isotropic, can be seen surrounding the gelled polymer phase.

These results show conclusively that even in the very short timescale involved in the elongation (<0.6 s), structural changes can be induced even in polymer solutions which are stable at rest. Because these structural changes are reversible, as shown by the redissolution, a weak gelation⁴ is observed.

Similar results have been observed with solutions of polymethyl methacrylate (PMMA) in dimethyl methane phosphonate (DMMP), polyethylene oxide (PEO) in water, polyacrylic acid (PAA) in ethylene glycol, polybutadiene (PBD) in dekaline, and others. In all cases, the polymers were of high molecular weight, see Table 1. Where polymers of lower molecular weight ($\leq 50,000$) were used, this phenomenon was not observed. The effect was most prominent in those solutions where the solubility parameters of solvent and solute were significantly different; for example, PMMA and DMMP have solubility parameters of 19.1 $\text{MPa}^{0.5}$ and 21.6 $\text{MPa}^{0.5}$ respectively. In this system, phase

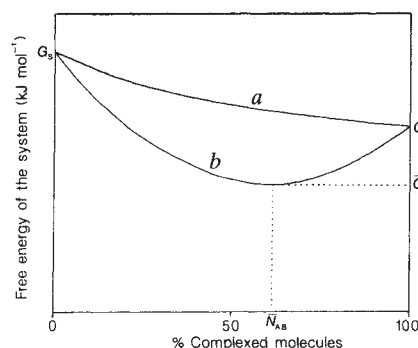


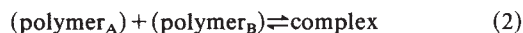
Fig. 4 Free-energy diagram for the complexation process. G_s is the total free energy of the separate constituents in the system. G_c is the total free energy of completely complexed constituents. N_{AB} is the percentage of complexed molecules. At equilibrium, the free energy is given by $\bar{G}$, when $N_{AB}\%$ molecules have complexed.

change was clearly observed. Therefore, both molecular weight and solvent compatibility are the crucial factors.

Gelation can be considered a special case of polymer-polymer complexation. Even in solutions with high intermolecular attraction between similar, or dissimilar, polymers, the formation of complexes is a comparatively long-term process⁵. In extensional flow, an additional factor must be present which is not only speeding up the rate of interaction, but in some cases, is enabling it to occur.

Mackley and Keller⁶ have shown that in a converging flow, which is predominantly extensional along the centre line, a very high degree of orientation is induced at high rates of elongation. We have also found⁷ that only a comparatively small number of long lifetime entanglements are required to cause this orientation (on average, two-thirds of the molecules have at most one entanglement when the solution changes from having a mainly viscous response to the imposed flow field, to having a mainly elastic response). Therefore, gel formation (complexation) is apparently accelerated in an extensional flow field. We can obtain support for this hypothesis from the thermodynamics of the process.

A typical free-energy diagram for the reversible reaction



is shown in Fig. 4. If $G_c > G_s$, then no complexation occurs. If the free energy falls monotonically from G_s to G_c , case *a*, then complete complexation takes place. This happens, for example, with solutions of PAA and poly vinyl pyrrolidone⁸. Another possibility is case *b*, where the minimum free energy, $\bar{G}$, at equilibrium occurs at some intermediate value of N_{AB} , denoted by $\bar{N}_{AB}$. If $\bar{N}_{AB}$ is small, then very little complexation is taking place. However, if $\bar{G}$ is increased, $\bar{N}_{AB}$ will also increase, until, when $\bar{G} = G_c$, total complexation takes place. The driving force for complexation is given by

$$\Delta G = \bar{G} - G_s \quad (3)$$

The change in free energy, ΔG , is made up of two components⁸,

$$\Delta G = \Delta G_1 + \Delta G_2 \quad (4)$$

ΔG_1 is the change in conformational free energy, given by

$$\Delta G_1 = \Delta H_1 - T\Delta S_1 \quad (5)$$

The change in enthalpy of conformation, ΔH_1 , is insignificant in comparison to the change in entropy of conformation, ΔS_1 .

ΔG_2 is the change in configurational free energy, given by,

$$\Delta G_2 = \Delta H_2 - T\Delta S_2 \quad (6)$$

ΔH_2 is zero for changes in configuration, whilst ΔS_2 , the change in configurational entropy, is related to the probability of the polymers being in solution or in complex⁸ and can be quite large.

In a flow field, the free energy of the solution will be increased.

Table 1 Molecular weights of the solutes used in the experiments

Solute	Molecular weight
FM9*	>5,000,000
PMMA	856,000
PEO	4,000,000
PAA	4,000,000
PBD	300,000

* A copolymer of tertiary butyl styrene and a monomer containing 'associating' groups, manufactured by ICI Plc.

In an extensional flow field, the uncoiling and aligning of the macromolecules will decrease ΔS_2 considerably. Hence, the driving force for complexation will increase. If molecules can be held close together by long lifetime entanglements, then complexation will proceed by a 'zipping up' process.

Therefore, it is suggested that the observed rise in elongational viscosity with strain is also due to increased intermolecular interaction between molecules in poor solvents, brought about by alignment in the elongational flow field.

Solutions of silk proteins show an inherent tendency to gel,

even at rest⁹. Given high intermolecular forces of attraction and the largely helical conformation of fibroin molecules, it would seem that on extrusion of an aqueous solution, followed by a high elongation rate, the phase change observed in solutions of synthetic polymers would occur. The immediate consequence of this would be that the complexed gel, being highly orientated, would have a much higher tensile strength than a homogeneous solution, enabling the weight of a spider to be supported almost instantaneously. Furthermore, the solvent-rich phase on the outside would be likely to evaporate at a very high rate, increasing the strength of the solidifying filament, although this has not been observed in the synthetic polymer systems studied.

Received 24 June; accepted 12 November 1986.

1. Trouton, F. T. *Proc. R. Soc. A* 77, 426-440 (1906).
2. Lodge, A. S. *Elastic Liquids* 229-231 (Academic, New York, 1964).
3. Ferguson, J., Hudson, N. E. & Warren, B. C. H. *J. Non-Newtonian Fluid Mech.* (in the press).
4. de Gennes, P.-G. *Scaling Concepts in Polymer Physics*, 134-136 (Cornell University Press, Ithaca, New York, 1979).
5. Challa, G. & Tan, Y. Y. *Pure Appl. Chem.* 53, 627-641 (1981).
6. Mackley, M. & Keller, A. *Phil. Trans. R. Soc. A* 278, 29-66 (1978).
7. Hudson, N. E. & Ferguson, J. *Trans. Soc. Rheol.* 20, 265-296 (1976).
8. Kabanov, V. A. & Papisov, I. M. *Polym. Sci. USSR* 21, 261-307 (1979).
9. Rundall, K. M. & Kenchington, W. A. *Rev. Entomol.* 16, 73-96 (1971).

²⁹Si magic-angle NMR spectroscopy of low-temperature ordered plagioclase feldspars

R. James Kirkpatrick*, Michael A. Carpenter†, Wang-Hong Yang* & Ben Montez*

* Mineral Physics Program, Department of Geology, University of Illinois, Urbana, Illinois 61801, USA

† Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK

Plagioclase feldspars are among the most abundant silicate minerals in the Earth's crust, and yet their subsolidus phase relations remain, perhaps, the most enigmatic¹. Their complex behaviour, which involves ordered (commensurate and incommensurate) structures, displacive transitions and three miscibility gaps, is generally attributed to variations in the ordering of aluminium and silicon between tetrahedral sites of the feldspar framework structure as a function of temperature and composition^{1,2}. The similarity in X-ray scattering factors of Al and Si atoms, however, has hindered conventional attempts to elucidate the exact structural relations. In the context of feldspar mineralogy, the recent advent of magic-angle sample spinning nuclear magnetic resonance spectroscopy (MAS NMR) has many implications, because it provides a powerful and direct method for determining the local distributions of Al and Si in solids³⁻⁸. The ²⁹Si-MAS NMR spectra of natural and synthetic plagioclases are equally complex, however, in containing many overlapping peaks and varying according to the chemistry and thermal history of each sample examined⁸⁻¹¹. We present here new spectra, obtained for a suite of carefully characterized ordered samples from metamorphic and slowly cooled igneous rocks, which shed light on: (1) the sensitivity of the feldspar framework to small amounts of Al/Si disorder; (2) the substitution mechanism of Si for Al in the solid solution from pure anorthite (CaAl₂Si₂O₈) towards albite (NaAlSi₃O₈); and (3) structural variations in the incommensurate ordered structure of intermediate compositions. In the wider context of silicate chemistry, NMR studies on feldspars provide insights into the subtle influences of continuous solid solution on crystal structures with changing composition. This is also the first attempt to examine the local structure of an incommensurate phase by NMR spectroscopy.

The spectra were obtained for homogeneous, ordered plagioclases which had been purified from natural metamorphic and slowly cooled igneous rocks and had almost all been characterized by transmission electron microscopy, electron microprobe analysis, X-ray powder diffraction and high-temperature solution calorimetry^{12,13}. They vary systematically in composition between albite and anorthite which enables us to follow spectral changes across the solid solution (Figs 1, 2). The positions of the best-resolved peaks at ~-111 p.p.m. (parts per 10⁶) (T1(0Al)) and -106 p.p.m. (T1(1Al))¹¹ do not vary significantly (~±1 p.p.m.) across almost the entire compositional range, and our assignments are therefore based on those made for ordered and heat-treated (that is, substantially disordered) Amelia albite¹¹. The limited variation with composition of the positions of many of the peaks suggests that the Na/Ca ratio has little direct influence on ²⁹Si-chemical shifts. Peaks in the range -81 to -92 p.p.m. must include those from Q⁴ (4Al) sites, as they feature most prominently in the ordered anorthite spectra (Figs 1, 3), whereas those between ~-92 and -112 p.p.m. must be due to Si on Q⁴(3Al), Q⁴(2Al), Q⁴(1Al) and Q⁴(0Al) sites, as in ordered and disordered albite.

The spectrum of pure metamorphic anorthite from Val Pasmada provides the key to behaviour at the anorthite end of the solid solution (Fig. 3a). In principle, a fully ordered anorthite (An₁₀₀) at room temperature should give a spectrum with eight peaks of equal intensity. On heating to ≥240 °C, pure anorthite undergoes a displacive transition, giving a symmetry change from Pī to Iī and a reduction in the number of non-equivalent Si-bearing Q⁴ (4Al) sites from 8 to 4 (ref. 14). The Val Pasmada anorthite spectrum has six clearly resolved peaks in three groups, each of two peaks. The intensity ratio of the three groups is 3:3:2, and the spectrum is well simulated by eight peaks with approximately equal intensities if two of the eight have, by chance, the same chemical shift (Table 1, Fig. 3a). Because of uncertainties in published refinements of the anorthite structure, however, the assignment of each peak to a specific silicon site cannot yet be made with confidence⁹.

Rather than obtain the Iī structure by heating (not possible at present with our spectrometers), it is possible to investigate the effect of a symmetry increase by following the addition of albite component in solid solution with anorthite, because this effectively depresses the temperature of the Iī to Pī transition¹⁵. If it has local Iī symmetry, a fully ordered Iī sample of bytownite should give a spectrum with a maximum of four Q⁴(4Al) peaks, and a range of non-Q⁴(4Al) peaks which will depend on the

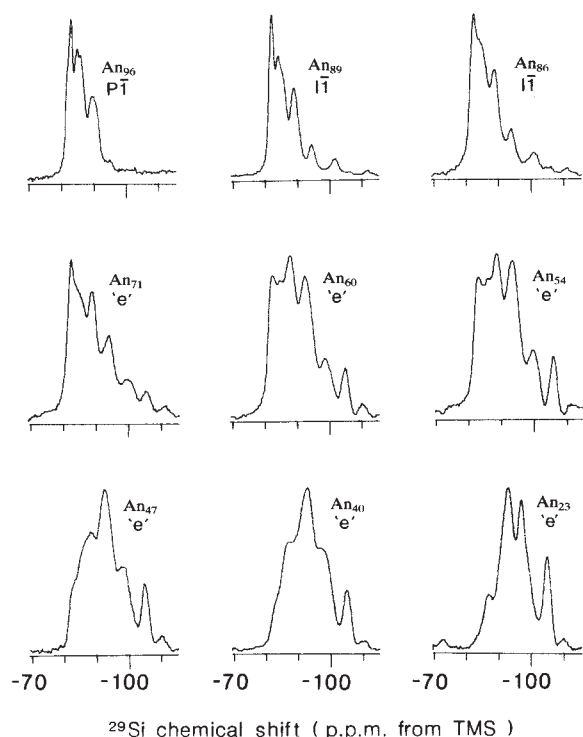


Fig. 1 ^{29}Si -MAS NMR spectra at 8.45 T of a suite of low-temperature, relatively well ordered P1, I1 and 'e' plagioclase feldspars. Except for the An_{71} sample, a few grains of which show evidence of exsolution on a fine scale¹², the samples were all free of intergrowth textures^{12,13}. The two samples not described previously¹² are 67783a, An_{46} (range An_{44-48}), from a granulitic anorthosite from the Uluguru mountains, Tanzania and 84340a, An_{55} (range An_{51-57}), from an amphibolite from the Pinnacles, Broken Hill, Australia (sample P6 of ref. 19); both give sharp 'e' reflections in electron diffraction patterns. The spectra were obtained as in ref. 21, except that a sample probe with a very stable rotor assembly was used (Doty Scientific).

sites occupied by the additional Si substituting for Al. The spectrum of metamorphic bytownite (An_{89}) in Fig. 3b shows just such a decrease in the number of peaks in the range -82 to -90 p.p.m. and has several additional peaks in the range -90 to -112 p.p.m. This interpretation is contradicted, however, by the presence of diffuse reflections at $h+k=2n$, $l=2n+1$ positions in electron diffraction patterns from the sample¹², and by the observation that the relative intensities of the $\text{Q}^4(4\text{Al})$ peaks in the NMR spectrum are not approximately equal; the local-symmetry within the crystals (over the range of a few unit cells) may really be P1. The NMR and diffraction observations can be reconciled if the positions of the Na and Ca atoms do not directly influence the ^{29}Si -chemical shifts, but are primarily responsible for the diffuse reflections. Alternatively, the feldspar framework is not distorted sufficiently from I1 symmetry to give resolvable peak splitting in the NMR spectrum, despite the effects of small P1 displacements of all the atoms. In this context, note that the spectra of synthetic and volcanic anorthites show significant peak broadening and overlap effects^{9,10} relative to metamorphic anorthite, even though they may have only a small amount of Al/Si disorder¹². The ^{29}Si -spectra thus provide information relating not only to different local tetrahedral site occupancies but also to the nature of framework distortions. There seems to be a rather sensitive relationship between these distortions and the degree of order.

Using the assignments for high albite of Yang *et al.*¹¹, non- $\text{Q}^4(4\text{Al})$ peaks in the An_{89} sample have been assigned primarily to T1(0Al) (-112 p.p.m.), T1(1Al) (-106 p.p.m.), T2(0Al) (-102 p.p.m.) and T1(3Al) (-95 p.p.m.); signal from other sites

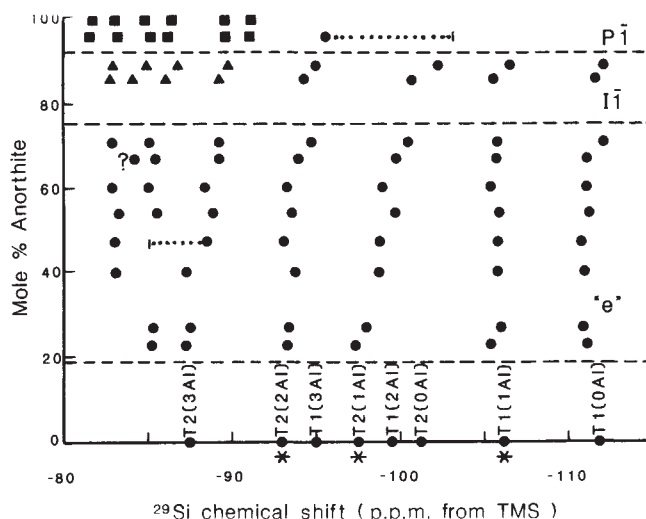


Fig. 2 Variation with bulk composition of the positions of the readily resolvable peaks and shoulders in the ^{29}Si -MAS NMR spectra of our plagioclase samples. ■, $\text{Q}^4(4\text{Al})$ peaks of P1 samples; ▲, $\text{Q}^4(4\text{Al})$ peaks of I1 samples; ●, other peaks; ... 1, broad band of intensity; peak assignments for pure albite are for low albite (indicated by *) and disordered albite¹¹. Note the lack of variation in the positions of the peaks for the T1(0Al) (-111 p.p.m.) and T1(1Al) (-106 p.p.m.) sites. The positions of the peak maxima indicate that in the -97 to -102 p.p.m. range, peaks for T2(1Al) sites dominate the spectra of low An-content samples, and peaks for T1(2Al) and T2(0Al) sites become relatively more intense with increasing An-content; in the -92 to -95 p.p.m. range, peaks for T2(2Al) sites dominate at low and intermediate An contents, and the peaks for T1(3Al) sites become relatively more intense at higher An contents. In the -87 to -91 p.p.m. range, peaks for T2(3Al) sites dominate at low An contents and the peaks for $\text{Q}^4(4\text{Al})$ sites become more intense with increasing An content. Precision and accuracy of the peak positions is $\sim \pm 0.5$ -0.8 p.p.m.

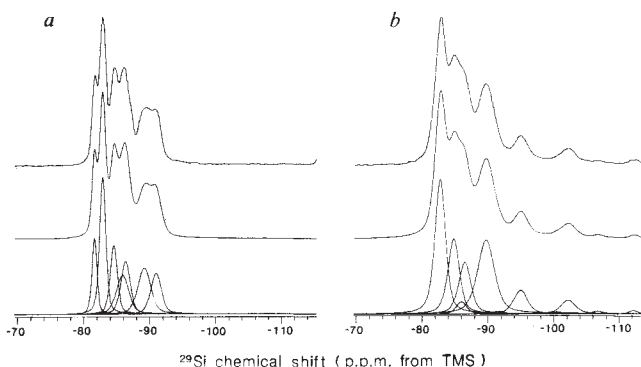


Fig. 3 *a*, Observed and simulated spectra and spectral components for the 8.45-T ^{29}Si -MAS NMR spectrum of Val Pasmada metamorphic anorthite (An_{100}); simulation obtained with the Nicolet curve fitting program CAP. *b*, Observed and simulated spectra and spectral components for the 8.45-T ^{29}Si -MAS NMR spectrum of the metamorphic Sittampundi bytownite (An_{89}); simulation obtained with the Nicolet curve fitting program CAP.

may be lost in the larger peaks. A first attempt at a simulation of the spectrum has been made including all these peaks and contributions from T2(3Al) (-88 p.p.m.), T1(2Al) (-100 p.p.m.), T2(1Al) (-97 p.p.m.) and T2(2Al) (-92 p.p.m.) sites (Table 1, Fig. 3b). Because of the strict alternation of Si and Al sites in fully ordered anorthite, each additional Si atom replacing an Al atom forms a $\text{Q}^4(0\text{Al})$ site and causes the neighbouring sites, of which there are four, to become $\text{Q}^4(3\text{Al})$.

Table 1 Peak positions, relative peak intensities, and peak assignments for the simulated ^{29}Si spectra of Val Pasmada An_{100} and Sittampundi An_{89}

Peak position	Relative intensity (%)	Site assignments
Val Pasmada An_{100}		
-81.7	11.5	$\text{Q}^4(4\text{Al})$
-83.0	25.1	$\text{Q}^4(4\text{Al})$
-84.8	11.9	$\text{Q}^4(4\text{Al})$
-85.5	12.7	$\text{Q}^4(4\text{Al})$
-86.3	12.4	$\text{Q}^4(4\text{Al})$
-89.6	13.6	$\text{Q}^4(4\text{Al})$
-91.1	12.8	$\text{Q}^4(4\text{Al})$
Sittampundi An_{89}		
-82.8	29.0	$\text{Q}^4(4\text{Al})$
-84.9	15.8	$\text{Q}^4(4\text{Al})$
-86.3	13.8	$\text{Q}^4(4\text{Al})$
-88.4	3.9	$\text{T}2(3\text{Al})$
-89.7	23.7	$\text{Q}^4(4\text{Al})$
-91.7	0.54	$\text{T}2(2\text{Al})$
-94.9	7.6	$\text{T}1(3\text{Al})$
-96.5	0.18	$\text{T}2(1\text{Al})$
-100.5	0.29	$\text{T}1(2\text{Al})$
-102.2	4.1	$\text{T}2(0\text{Al})$
-106.5	0.55	$\text{T}1(1\text{Al})$
-112.0	0.67	$\text{T}1(0\text{Al})$

An_{100} simulated assuming $\text{P}\bar{1}$ local structure (eight Si sites); An_{89} simulated assuming $\text{I}\bar{1}$ local structure (four Si sites).

If such a substitution mechanism extends to An_{89} , 5.2% of the Si sites should be $\text{Q}^4(0\text{Al})$, which compares quite closely with the total intensity (4.8%) of the two peaks assigned to this type of site in the simulation. The apparent preponderance of $\text{T}1(3\text{Al})$ sites (-95 p.p.m.) relative to $\text{T}2(3\text{Al})$ (-88 p.p.m.) and of $\text{T}2(0\text{Al})$ (-102 p.p.m.) relative to $\text{T}1(0\text{Al})$ (-112 p.p.m.) implies also that the excess Si substitutes preferentially for Al on T2 sites rather than on T1 sites. (Each T2 site in the feldspar framework shares oxygens with three T1 sites and one T2 site, whereas each T1 site shares oxygens with three T2 sites and one T1 site¹⁴.) Recent structure refinements of bytownites are consistent with this preference of Si for T2 over T1 sites^{16,17}. If the excess Si prefers T2 to T1 by five to one but is otherwise randomly distributed, the expected concentration of $\text{T}1(1\text{Al})$ sites (at -106 p.p.m.) is 0.2%, compared with the observed concentration of 0.55% in the simulation. This observation is consistent with some clustering of the sites occupied by the excess Si.

The most striking feature of the spectra of the plagioclase crystals with incommensurate (or 'e') ordered structures is the narrowness of their most prominent peaks (Fig. 1). These spectra

must be made up of a number of overlapping peaks, but the presence of some relatively intense and narrow peaks implies that there are many Si sites in the crystals with identical environments. Presumably, therefore, the incommensurate structure possesses a rather high degree of Al/Si order. Certainly, none of the spectra we have seen for disordered plagioclases of any composition has such well-resolved peaks. Other experimental studies have suggested that there may be significant differences between 'e' structures in the composition ranges An_{50} - An_{70} (designated 'e₁') and An_{20} - An_{50} (designated 'e₂')^{1,2,12,13,18,19}. The MAS NMR spectra appear to reflect these differences (Figs 1, 2). Spectra from the 'e₁' samples (An_{54} , An_{60} , An_{71}) all contain well resolved peaks or shoulders at ~-83, -85 and -89 p.p.m., presumably due to $\text{Q}^4(4\text{Al})$ sites of similar character to those in An_{89} . With increasing Ab content in this composition range, the relative intensities of the $\text{Q}^4(4\text{Al})$ sites decrease (except for the peak at -89 p.p.m., which overlaps with the $\text{T}2(3\text{Al})$ peak) and the total relative intensity of the peaks for non- $\text{Q}^4(4\text{Al})$ sites increases. The non- $\text{Q}^4(4\text{Al})$ sites present, however, are not entirely those of either ordered low albite or disordered high albite. In particular, there is little intensity for $\text{T}2(1\text{Al})$ sites at ~-97 p.p.m., a dominant peak in albite spectra^{9-11,20-22}. The spectra of 'e₂' plagioclases with compositions from An_{23} to An_{40} (and possibly the An_{47} sample) contain significant intensity for the three peaks of low albite at -93, -97 and -105 p.p.m. but do not contain well-defined peaks for $\text{Q}^4(4\text{Al})$ sites. These spectra will provide a basis for testing site-occupancy models of the incommensurate plagioclase structures, but are not easy to interpret directly. For the present, they merely support the view that slab models involving only layers of albite and anorthite cannot be correct^{1,9,12,13,23,24}, that the incommensurate structure really is rather well ordered, and that the distinction between 'e₁' and 'e₂' incommensurate types may be structurally significant.

We conclude that the ^{29}Si -MAS NMR spectra for plagioclase feldspars contain a wealth of quantitative structural information. We are attempting to unravel some of the details by systematic studies of specially selected and characterized samples and by simulation of the spectra with different structural models of 'e' plagioclases. Distortions of the feldspar framework are particularly sensitive to relatively small variations in Al/Si order and the NMR technique emphasizes this sensitivity. In contrast, we suggest that the peak positions are relatively insensitive to variations in Na/Ca content. This type of approach to silicate chemistry gives insights into details both of the structures of individual stoichiometric phases and of the local structure variations within crystals which are due to the formation of continuous solid solutions.

This research was supported by NSF grant EAR84-08421 to R.J.K. and NERC grant no. GR3/5547 to M.A.C. We also thank Dr E. Salje and Dr A. Putnis for their critical comments. Cambridge Earth Sciences contribution ES826.

Received 24 April; accepted 27 October 1986.

- Smith, J. V. in *Feldspars and Feldspathoids* (ed. Brown, W. L.) 55-94 (NATO ASI series C, 137, Reidel, Dordrecht, 1984).
- Ribbe, P. H. *Rev. Miner.* **2**, 1-55 (1983).
- Fyfe, C. A., Thomas, J. M., Klinowski, J. & Gobbi, G. C. *Angew. Chem. Int. Edn Engl.* **22**, 259-336 (1983).
- Oldfield, E. & Kirkpatrick, R. J. *Science* **227**, 1537-1544 (1985).
- Kirkpatrick, R. J., Smith, K. A., Schramm, S., Turner, G. & Yang, W.-H. *A. Rev. Earth planet. Sci.* **13**, 29-47 (1985).
- Putnis, A., Fyfe, C. A. & Gobbi, G. C. *Phys. Chem. Miner.* **12**, 211-216 (1985).
- Putnis, A. & Angel, R. J. *Phys. Chem. Miner.* **12**, 217-222 (1985).
- Higgins, J. B. & Woessner, D. E. *Eos* **63**, 1139 (1982).
- Kirkpatrick, R. J., Kinsey, R. A., Smith, K. A., Henderson, D. M. & Oldfield, E. *Am. Miner.* **70**, 106-123 (1985).
- Sherriff, B. L. & Hartman, J. S. *Can. Miner.* **23**, 205-212 (1985).
- Yang, W.-H., Kirkpatrick, R. J. & Henderson, D. M. *Am. Miner.* **71**, 712-726 (1986).
- Carpenter, M. A., McConnell, J. D. C. & Navrotsky, A. *Geochim. cosmochim. Acta* **49**, 947-966 (1985).

- Carpenter, M. A. *Phys. Chem. Miner.* **13**, 119-139 (1986).
- Ribbe, P. H. in *Feldspars and Feldspathoids* (ed. Brown, W. L.) 1-54 (NATO ASI series C, 137, Reidel, Dordrecht, 1984).
- Adhart, W., Frey, F. & Jagodzinski, H. *Acta crystallogr.* **A36**, 461-470 (1980).
- Chiari, G., Benna, P. & Bruno, E. *Z. Kristallogr.* **169**, 35-49 (1984).
- Wenk, H.-R. & Kroll, H. *Bull. Miner.* **107**, 467-487 (1984).
- McConnell, J. D. C. in *The Feldspars* (eds MacKenzie, W. S. & Zussman, J.) 460-477 (Manchester University Press, 1974).
- Slimming, E. H. *Am. Miner.* **61**, 54-59 (1976).
- Lippmaa, E., Magi, M., Samoson, A., Engelhardt, G. & Grimmer, A.-R. *J. Am. chem. Soc.* **102**, 4889-4893 (1980).
- Smith, K. A., Kirkpatrick, R. J., Oldfield, E. & Henderson, D. M. *Am. Miner.* **68**, 1206-1215 (1983).
- Smith, J. V., Blackwell, C. S. & Hovis, G. L. *Nature* **309**, 140-142 (1984).
- Kitamura, M. & Morimoto, N. in *Feldspars and Feldspathoids* (ed. Brown, W. L.) 95-119 (NATO ASI series C, 137, Reidel, Dordrecht, 1984).
- Horst, W. *Fortschr. Miner.* **62**, 107-154 (1984).

Structure of the Kenya rift from seismic refraction

KRISP working group*

The Kenya rift forms part of the East African rift system, a tensional feature sometimes regarded as representing an early stage of continental break-up^{1,2}. High-resolution seismic refraction data are required to provide constraints on models of the deep structure and processes. A 300-km seismic refraction line along the axis of the southern portion of the rift shows variations of up to 4 km in the depth to the basement. A 7.6 km s^{-1} layer at $\sim 35 \text{ km}$ depth beneath the rift is overlain by a 10-km-thick lens which appears to include anomalous low-velocity material. This lens is interpreted to represent the base of the crust and thickens beneath the Kenya dome. There is no clear evidence for a wide high-velocity intrusive zone in the upper crust within the rift.

A long-range explosion seismology experiment was carried out in August 1985 as part of KRISP 85 (Kenya Rift International Seismic Project 85) by a group of institutes from the United Kingdom, United States, West Germany, Switzerland, Kenya and Zimbabwe. The major aims of this experiment were (1) to determine the velocity-depth structure beneath this portion of the rift and (2) to test shot efficiency, recording station characteristics, and transmission along and across the rift and its boundaries. This information is required for the planning of a series of further experiments in 1988 and beyond.

The experiment consisted of two seismic refraction lines (Fig. 1). Along the rift, a N-S line was completed with shots at Lake Baringo (BAR), Chepkererat (CHE), Solai (SOL), Elmenteita (ELM), Naivasha (NAI), Susua (SUS), and Magadi (MAG) recorded by 42 three-component stations at 3.5 km intervals between CHE and SUS. Across the rift an E-W line was completed with shots at Ewaso Ngiro (EWA), Ntulelei (NTU), Susua (SUS), Mount Margaret (MAR), and Makuyu (MAK) into the same stations deployed at 1.25 km intervals between NTU and MAR. On both lines, 8 additional stations were used to fill gaps between the end and off-end shots at larger intervals. The explosion programme was followed by three earthquake recording experiments between September and December 1985.

In the vicinity of the seismic experiment, the rift is 50–70 km wide (Fig. 1) and is flanked by shoulders 1–2 km above the rift floor, the major boundary faults having throws of up to 3–4 km. It is covered by volcanics and sediments^{2–4} to a depth of $\geq 2 \text{ km}$ and is dissected by subparallel, north-south trending normal faults. Precambrian rocks are exposed in rift boundary faults to the north (Elgeyo escarpment) and south (Nguruman escarpment) of the central section, and in the Saimo fault block within

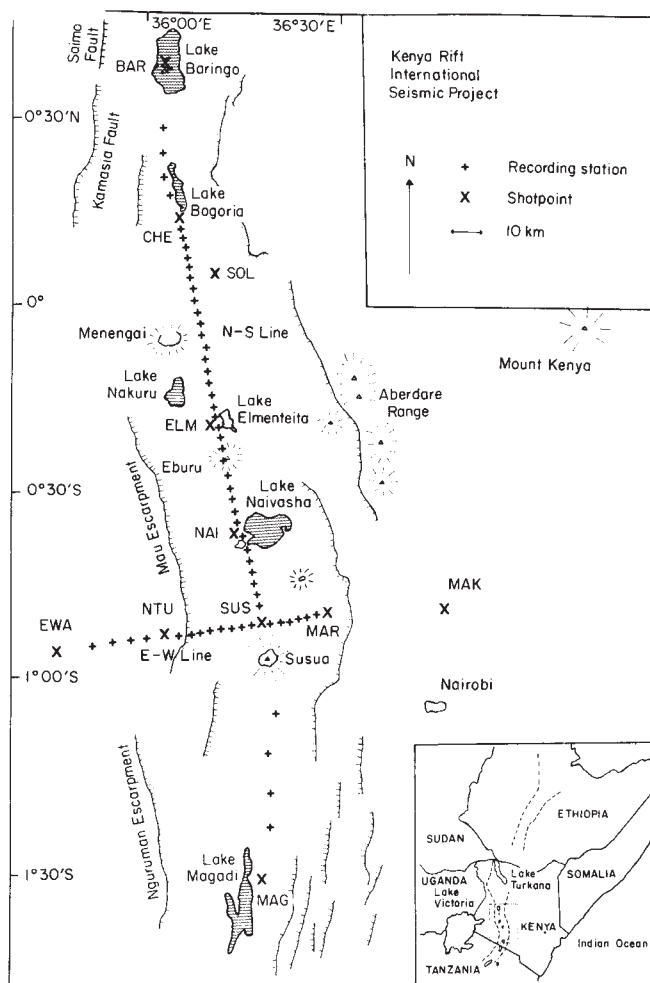


Fig. 1 Location of the KRISP 85 seismic experiment showing the lines and shot points. Recording station locations for the east-west line are diagrammatic, as the spacing is too small to be shown on this scale.

the rift to the east of the Elgeyo. The oldest rift structures are mid-Tertiary and immediately pre-date the rift volcanism³ which began in the Miocene about 23 Myr ago² and has continued to the present as is evident by the abundant volcanic craters (Fig. 1).

Geophysical models based on gravity^{3,5,6} and teleseismic data⁷ suggest first, lithospheric attenuation, enabling anomalously low-velocity, low-density asthenospheric material to reach to the base of the crust beneath the Rift Valley and thus replace a wide mantle zone. It is suggested as a mechanism providing isostatic compensation for the raised topography of the Kenya dome⁸; and second, the models suggest the presence of a relatively high-velocity, high-density intrusive body penetrating the crust to shallow levels from the anomalous mantle zone beneath the axis of the rift.

The location of our north-south line was chosen to overlie this 'axial intrusion' as closely as possible. A seismic refraction line completed in 1968 along the rift axis to the north of Lake Bogoria^{9,10} also suggests asthenospheric penetration of the lithosphere to within 15–20 km of the surface. A shorter refraction line completed in 1975 west of Lake Baringo⁴ failed to give any direct evidence for the existence of the axial intrusion, although the presence of anomalously high density material in the upper crust was implied by the combined interpretation of the refraction and the gravity data.

Recording stations were positioned by conventional and aerial survey techniques to an accuracy of 20 m. Just under half of the data were recorded digitally on US equipment (Kinometrics

* M. A. Khan, P. K. H. Maguire, W. Henry, M. Higham (Geology Department, University of Leicester, UK). G. R. Keller, S. Harder (Geology Department, University of Texas, El Paso, USA). C. Prodehl, J. Mechie, W. Kaminski, Geophysikalisches Institut, University of Karlsruhe, FRG). C. J. Swain (Department of Physics, University of Zimbabwe, Harare, Zimbabwe). J. R. Bowman, (School of Earth Sciences, Australian National University). L. W. Braille (Department of Geosciences, University of Purdue, USA). J. Davies (Continental Ore, Nairobi, Kenya). D. H. Griffiths, R. F. King (Department of Geological Sciences, University of Birmingham, UK). B. Guggisberg (Geophysikalisches Institut, ETH, Zurich, Switzerland). R. P. Meyer, V. Green (Department of Geophysics, University of Wisconsin-Madison, USA). A. E. Mussett (Department of Geology, University of Liverpool, UK). M. Mwangi (Olkaria Geothermal Project, Kenya Power and Lighting Company, Naivasha, Kenya). F. Mwangi (Ministry of Water Development, Nairobi, Kenya). K. H. Olsen (Los Alamos National Laboratory, New Mexico, USA). J. Patel (Department of Physics, University of Nairobi, Kenya). G. A. Thompson (Department of Geophysics, Stanford University, USA). J. Wohlenberg (Department of Geology, University of Aachen, FRG).

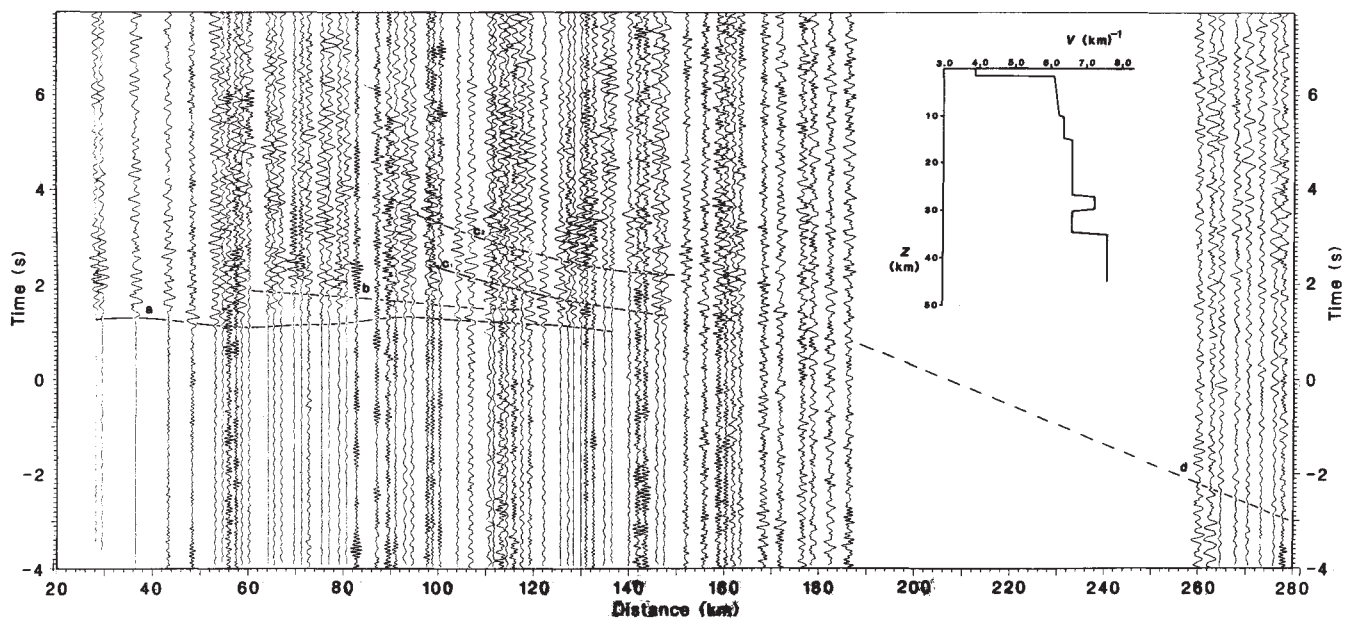


Fig. 2 Record section from shots BA1 and BA2 with phase correlations and velocity-depth function. Phase d is continued across the data gap for the reader's convenience. (Reducing velocity = 6.0 km s^{-1}).

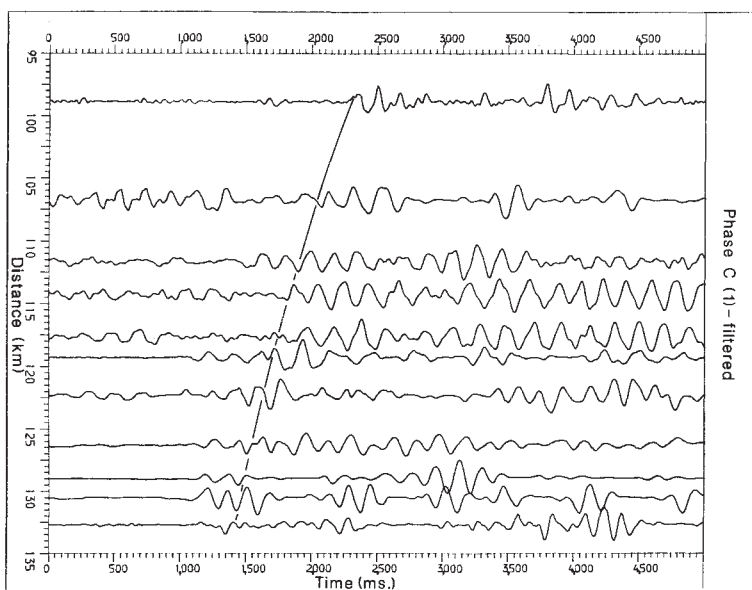


Fig. 3 Enhancement of phase c_1 by polarization filtering (reducing velocity = 6.0 km s^{-1}).

DR100 and University of Wisconsin recorders) and the rest on frequency-modulated analogue systems (European MARS and UK Geostores). At each station an Omega-controlled time signal was recorded in addition to the three seismic components. The analogue data were subsequently digitized and merged with the digitally recorded data at 100 samples per second.

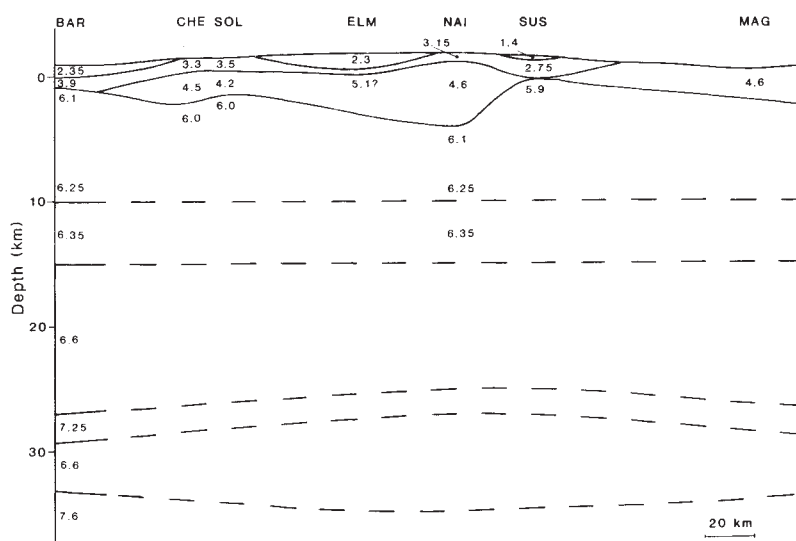
The lake shots were efficient. For example, as can be seen on the section shown on Fig. 2, a 1-tonne shot in Lake Baringo provided useful energy to 280 km. In contrast, the borehole shots were inefficient—a 1-tonne shot distributed between six holes at 60 m depth at MAG did not provide useful energy beyond 100 km. Two factors contributed to this—(1) the shot holes did not reach the water table and (2) the recording stations were largely on pyroclastics which were associated with high noise levels. This was particularly serious on the east-west line as no lakes were available. For this reason, the data for this line are not included in the analysis at this time. However, this experience of borehole shooting in Kenya has provided the

information for the planning of future experiments which was one of the objects of the programme.

Frequency filtering, and 'REMODE'¹¹-type polarization filtering techniques were applied to the data to improve the signal to noise ratio. A P-wave record section of the data for the two shots at Baringo is shown in Fig. 2 together with our subjective phase correlations. The traces shown have been bandpass-filtered between 1 and 20 Hz and normalized. Figure 3 shows a portion of the same data after the application of a polarization filter. Fuller presentations of the data will be made elsewhere¹⁹.

Analysis of phases observed up to a few tens of kilometres from shots on our north-south line and previous experiments provides evidence that the rift infill has velocities ranging from 1.4 to 4.6 km s^{-1} and thicknesses ranging from 2 km beneath BAR and SUS to 6 km beneath NAI (Fig. 4). Phase a (P_g) can be recognized as a first arrival on the record section up to about 130 km with an apparent velocity of about 6.0 km s^{-1} . From the reverse shot, NAI, phase a has a higher apparent velocity and

Fig. 4 Proposed velocity-depth structure along the axis of the rift from Lake Baringo to Lake Magadi (velocities shown in km s^{-1}).



thus the two-dimensional model shows the uppermost basement dipping southwards between Elmenteita and Naivasha with velocities of about 6.1 km s^{-1} increasing to 6.25 km s^{-1} at 10 km depth. Phase b can be observed with high amplitudes between distances of 70 and 120 km. It is interpreted as a reflection from an intercrustal discontinuity at about 15 km depth. Between 15 and 27 km the velocity is about 6.6 km s^{-1} . Phase c_1 can be recognized, especially on the polarization filtered record section (Fig. 3) produced from selected three-component stations at distances between 100 and 130 km. It is interpreted as a reflected phase from a discontinuity at about 27 km depth below which a velocity of about 7.25 km s^{-1} occurs. Phase c_2 , interpreted as a reflection from a discontinuity at about 35 km depth, can be observed beyond about 100 km. In order to produce the large amplitudes for this phase at distances around 100 km, a layer with a low velocity of about 6.6 km s^{-1} is required between 30 and 35 km depth. On the record section, a few first arrivals occur in the 260–280 km range at -2 to -3 seconds reduced time. Because of their high amplitudes these arrivals are interpreted as diving waves (phase d) within a layer of velocity 7.6 km s^{-1} at its top. The main features of the velocity-depth function are shown in Fig. 2. Travel-time and amplitude analysis of the phases in the 50–150 km range suggest that the velocities shown are maximum estimates. Decreases of 0.1 and 0.25 km s^{-1} in the 6.6 and 7.25 km s^{-1} would provide acceptable fits to the data. However, these decreases will affect the depth estimates to the 7.25 and 7.6 km s^{-1} layers by less than 2 km.

The traditional definition of the Moho is not easily applied to this area of unusual velocity structure. The widely accepted definition is "that level in the earth where the compressional wave velocity increases rapidly or discontinuously to a value between 7.6 and 8.6 km s^{-1} " (ref. 12). However, from deep seismic sounding experiments in Afar and Ethiopia, Berckhemer *et al.*¹³ and Ruegg¹⁴ attribute velocities as low as 6.8 km s^{-1} to the upper mantle. Regardless of whether we call the 7.2 or 7.6 km s^{-1} layer the uppermost mantle, if we assume that the interpretation of the 1968 refraction data⁹ is correct, then our results demonstrate a thickening of the crust along the rift axis towards the centre of the Kenya dome. This thickening is consistent with the regional gravity data^{15,16} which shows that gravity values decrease from greater than -100 mgal around southern Lake Turkana through about -160 mgal near Lake Baringo to about -200 mgal near Lake Naivasha.

Banks and Swain⁸ studied the isostatic response of an 800 km square area surrounding the Kenya rift and favoured the idea that the compensation of the Kenya topographic dome is because of anomalous, low density material in the upper mantle. However, the results presented here suggest the compensation

is due, at least in part, to crustal thickening.

It has long been suggested that an 'axial intrusion' rising to very high levels in the crust is responsible for the gravity high along the axis of the rift^{3,5}. The basement velocities (6.1 km s^{-1}) measured by our experiment are rather low to be associated with such a dense intrusion so it is natural to ask if some other feature could account for the gravity high. Fairhead¹⁷ suggested a combination of volcanic infill and a relatively narrow dyke to model a profile along our east-west line. His interpretation mainly differed from Searle's⁵ in its choice of regional field which (unlike Searle's) was assumed to follow closely the observed Bouguer anomaly over exposed sections of the basement both west and east of the rift. This regional field yields positive residuals along the rift axis ($+20$ to 30 mgal) which cannot therefore be due to sub-volcanic basement topography within the rift. The source must lie within the basement. The 6.1 km s^{-1} could represent intrusive material, since lower basement velocities have been measured off the axis⁴, and velocity increases with depth (possibly indicative of dyke injection).

Savage and Long⁷ suggest that the short-wavelength pattern of teleseismic P-wave delays between Kaptagat and Nairobi is also primarily due to the existence of the high-velocity, axial intrusion rising to a level of 10–20 km within the rift. Since our results do not support the existence of such a feature continuously along the line, we suggest that variations in crustal structure (in particular basement topography) both across and along the rift axis make an important contribution to the reported pattern of P-wave delays. Further work on the proposed axial intrusion is clearly necessary as it has an important bearing in the low angle detachment models now gaining favour¹⁸.

In Fig. 4 we present a preliminary model synthesizing the results of the KRISP85 experiment and relevant data from previous studies. The model demonstrates that there is considerable basement topographic relief. The most important new feature of the velocity-depth model is the presence of a 10-km-thick lens at a depth of 25 km of material with a velocity of 7.2 km s^{-1} at its top decreasing by about 10% at the base. The study also suggests that there are significant variations in the crustal structure along the axis of the rift. Both of these features require further investigation.

We thank the following for their participation in the programme: U. Achauer, R. Aster, P. Bowry, M. Brooks, R. Catchings, B. Chege, P. Cooke, P. Coward, J. Doody, S. Gitonga, E. Goodwin, A. Haston, H. Hoffman, S. Holbrook, C. Jarchow, L. Johnson, R. Khan, Y. Khan, K. M. King, S. Kimani, N. Laffoley, H. Laske, T. Le Bas, R. Livesey, J. McCarthy, C. Mann, R. Mobbs, P. Morgan, D. Mutea, E. Mwango, M. Nganga, J. K. Nzuki, N. Ochmann, D. Okaya, S. Onancha, E. Smith,

R. Stellrecht, J. Swainston, W. Teasdale and H. Wachira. William Bosworth made helpful suggestions. Special thanks are due to the Kenya Government and its officials who made the work possible. The University of Nairobi, the Department of Mines and Geology, the Ministry of Water Development, Continental Ore, the Remote Sensing Centre, the Survey of Kenya, Kenya Oilfield Services, Inside Africa Safaris and Twiga Chemical Industries all made vital contributions to the programme. The work was financed by the NERC in the UK, NSF and Amoco in the US, the DFG in West German, the ETH in Zurich, the University of Leicester, and the Kenya Power and Lighting Company.

Received 13 June; accepted 20 November 1986.

1. Molner, P. & Oliver, J. *J. geophys. Res.* **74**, 2648-2682 (1969).
2. Baker, B. H., Mohr, P. A. & Williams, L. A. *J. Geol. Soc. Am. Spec. Paper No. 136* (1972).
3. Baker, B. H. & Wohlenberg, J. *Nature* **229**, 538-542 (1971).
4. Swain, C. J., Khan, M. A., Wilton, T. J., Maguire, P. K. H. & Griffiths, D. H. *J. Geol. Soc. Lond.* **138**, 93-102 (1981).
5. Searle, R. C. *Geophysical J.* **21**, 13-31 (1970).
6. Khan, M. A. & Mansfield, J. *Nature* **229**, 72-75 (1971).
7. Savage, J. E. G. & Long, R. E. *Geophys. J.* **82**, 461-477 (1985).
8. Banks, R. J. & Swain, C. J. *Proc. R. Soc. A* **364**, 331-352 (1978).
9. Griffiths, D. H., King, R. F., Khan, M. A. & Blundell, D. J. *Nature* **229**, 69-71 (1971).
10. Griffiths, D. H. *Tectonophysics* **15**, 151-156 (1972).
11. Kanasevich, E. R. in *Time Sequence Analysis in Geophysics*, 301-304 (The University of Alberta Press, 1975).
12. Steinhardt, J. S. in *International Dictionary of Geophysics* (ed. Runcorn, S. K.) (Pergamon, Oxford, 1966).
13. Berckhemer, H. et al. in *IUGG Report 14, Afar Depression of Ethiopia* (eds Pilger, A. & Rodler, A.) 89-108 (Schweizerbart'sche, Stuttgart, 1974).
14. Ruegg, J. G. in *IUGG Report 16, Afar—Between Continental and Oceanic Rifting* (eds Pilger, A. & Rosler, A.) 113-120 (Schweizerbart'sche, Stuttgart, 1974).
15. Swain, C. J. & Khan, M. A. in *Geological Background to Fossil Man* (ed. Bishop, W. W.) 1-13 (Scottish Academic Press, Glasgow, 1978).
16. Swain, C. J. & Khan, M. A. *Geophys. J.* **53**, 427-429 (1978).
17. Fairhead, J. D. *Tectonophysics* **30**, 269-298 (1976).
18. Bosworth, W., Lambiase, J. & Keisler, R. *EOS* **67**, 577-583 (1986).
19. Keller, G. R., Khan, M. A. & Prodehl, C. *EOS* **67**, 1103 (1986).

Late-Holocene upper timberline variation in the southern Sierra Nevada

Louis A. Scuderi

Isotope Laboratory, Institute of Geophysics and Planetary Physics, University of California, Los Angeles 90024, USA

Alpine timberline has been shown to be a definite tension zone between trees and climate where trees invade higher ground during favourable climatic periods and retreat during periods of deteriorating conditions¹. These movements are indicative of an upper forest limit sensitively adjusted to climatic variation. Here I present absolutely dated dendrochronological records from living trees and standing snags and radiometric dates from relict logs of foxtail pine at Cirque Peak, southern Sierra Nevada, California, providing the first detailed record of temperature-induced mid- to late-Holocene fluctuations in Sierran timberline. Marked declines in timberline correspond to lichen-dated cold periods and/or glacial advance.

Chronologies of variation in altitude of upper timberline reflecting late-Holocene climatic change have been established for several alpine areas in western North America²⁻⁶. These timberlines are characterized by zones of standing snags at and just above present timberline, rooted and fallen snags above this zone, and small deeply weathered remnants at the highest limit of relict wood. Similar conditions exist in several areas of the Sierra Nevada, with relict timberlines being especially well preserved in the drier southern portion of the range.

Foxtail pine (*P. balfouriana* Grev. and Balf.) is found at the timberline on Cirque Peak in the southern Sierra Nevada (Fig. 1). This species is closely related to the Great Basin bristle-

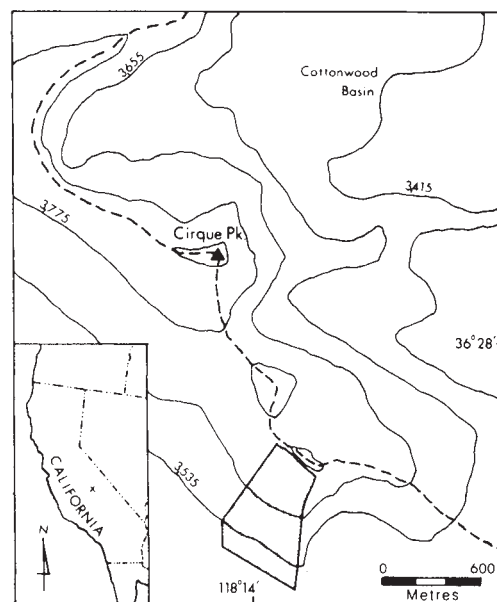


Fig. 1 Topographic map showing the Cirque Peak upper timberline site (shaded) and its relationship to Cottonwood Basin. Elevations in metres.

cone pine (*P. longaeva* Bailey) found 60 km northeast in the White and Inyo Mountains. Foxtail pine is characterized by high temperature sensitivity at its upper growth limit, excellent cross-dating qualities, and a relatively long lifespan. Living and relict individuals with lifespans exceeding 1,200 yr have been found at Cirque Peak⁷.

Computer-aided cross-dating⁸ of relict wood above the present timberline permitted the construction of an absolutely dated 3,031-yr chronology of ring-width variation for the Cirque Peak site⁷. Samples from 6,300-3,100 yr BP were dated radiometrically using the earliest portion of the remnant. Sample elevation was recorded at the time of collection and used in conjunction with dendrochronological and calibrated⁹ radiometric sample ages to produce a chronology variation in timberline elevation (Fig. 2).

Timberline was defined as the upper limit at which trees were continually established, a position now indicated by a sharp transition from upright trees to isolated outlier krummholz forms. The minimum timberline position for each time period was inferred from five to six cross-dated samples of upright, non-krummholz wood.

The oldest and highest sample located is a small weathered remnant 68 m above the present timberline, radiocarbon dated⁹ at $6,300 \pm 300$ yr BP (UCLA 2418-A). The wide growth rings of this and other cross-dated remnants at this elevation are indicative of warm, highly favourable growing conditions during the Altithermal. A date⁹ of $3,530 \pm 50$ yr BP (UCLA 2463-C) on a sample 65 m above the present timberline indicates that maximum levels were maintained until at least 3,500 yr BP.

The possibility of a higher Altithermal timberline, similar to that documented in the White Mountains^{2,3}, was explored at length. Slopes with suitable environmental factors, and lacking physical characteristics that would inhibit upward movement of the timberline, extend nearly to the crest of the range (240 m above the present timberline). But so far no relict wood has been found higher than 70 m above present levels for up to 1 km from the study site. In the light of the observed excellent sample preservation and lack of fire scars it appears unlikely that the timberline at Cirque Peak was much higher than 70 m above its present level at any time since 6,300 yr BP. Relict timberlines on other local peaks suggest that this may represent the maximum Altithermal timberline elevation for the region.

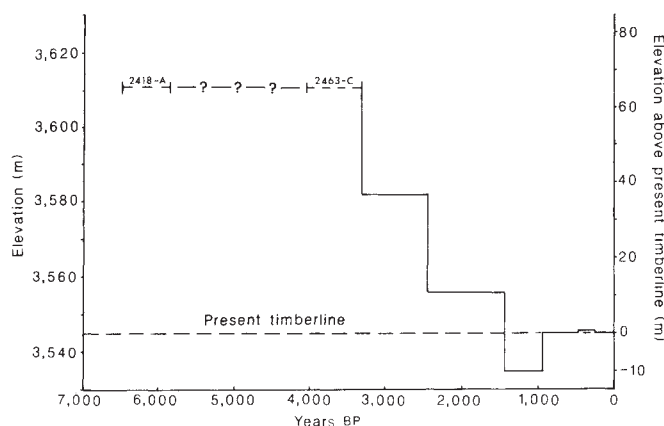


Fig. 2 Elevation of mid- to late-Holocene timberline at Cirque Peak. Sample numbers of UCLA radiocarbon dates are shown at the highest level of remnant wood 65–68 m above the present timberline.

High Altithermal levels were followed by a rapid decline in timberline of 30 m between 3,400 and 3,200 yr BP. Extrapolation of sample ages to pith dates shows that a large number of trees were established 30–37 m above present levels between 3,200 and 2,800 yr BP. This suggests that the timberline stabilized ~35 m above present levels by 3,200 yr BP and possibly several hundred years earlier.

Many trees, especially those at elevations greater than 20 m above the present timberline, died between 2,500 and 2,300 yr BP. A marked decrease in ring widths in all samples at 2,400 yr BP is suggestive of colder conditions leading to decreased growth and regenerative potential. A study of late-Holocene glaciation in the adjacent Cottonwood Basin⁷ indicates that this period was one of marked glacial advance, correlating to the earliest Recess Peak glaciation in the Sierra Nevada¹⁰. The timberline fell ~25 m to 12 m above the present timberline at this time, where it stayed for ~900 yr. Between 1,400 and 1,300 yr BP the timberline fell rapidly to 10 m below its present level. This decrease coincides with a sharp decline in ring widths for all trees from 1,400 to 1,000 yr BP. Similar small ring widths are also found between 1,300 and 1,050 yr BP at Sheep Mountain in the White Mountains¹¹. Lichenometric dating of deposits in the adjacent Cottonwood Basin⁷ and the Mammoth Lakes region further north¹⁰ indicates possible increased glacial activity during this interval.

The period 950–850 yr BP was one of increasing tree establishment during which the timberline rose 10 m to its present level. A few trees were established marginally above present levels at 450 yr BP, suggesting that the timberline may have peaked at this time. The onset of Matthes 'little ice age' advances at 350 yr BP does not appear to have significantly affected timberline position. The establishment of some small erect seedlings 5–10 m above the present timberline since 100 yr BP suggests a renewed rise in timberline due to recent warmer conditions^{2,3} or possible increases in atmospheric carbon dioxide concentration¹². It remains to be seen if these trees will survive to maturity or if they represent short-term variations in timberline that are not resolvable in the context of the coarser record of timberline.

Comparison with timberline records from the White and Rocky Mountains shows several points of similarity as well as some divergence between the records. Of note is the net difference in timberline elevation between present and past levels, especially from 6,300 to 3,500 yr BP. The Sierra Nevada and White Mountain timberlines were both at maximum elevation during this period, but the Cirque Peak timberline was only 70 m above its present level whereas the Campito and Sheep Mountain timberlines were 110 and 150 m respectively above their present levels². Pollen evidence of coarser resolution from

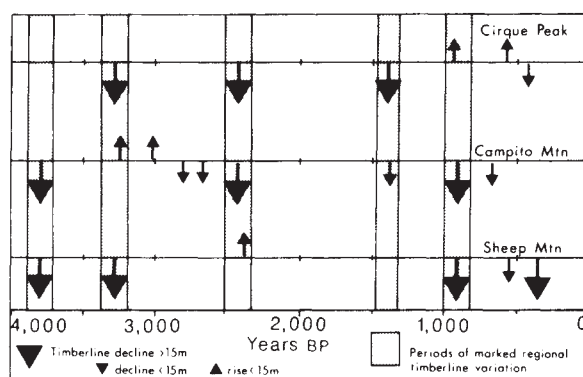


Fig. 3 Major timberline elevation changes at Cirque Peak and neighbouring White Mountain sites. Periods of synchronous timberline variation are indicated at approximately 3,800, 3,400, 2,400, 1,400 and 900 yr BP, indicating possible regional climatic forcing of the upper timberline position.

the Rocky Mountains indicates that timberline was then 70–100 m above present levels.

The regular nature of the current Cirque Peak timberline, which is characterized by a sharp transition from upright trees to treeless conditions and a smooth drop in elevation from southwest to northwest slopes, suggests that this timberline is a true 'climatic' timberline. The upper limit of remnant wood parallels the current timberline and also shows a similar distinct boundary. The sharply defined timberline and apparent lack of site specific microclimatic and lithological forcing of timberline position is in marked contrast to other more variable upper timberline limits in western North America. This regularity, as well as differences in defining timberline position, may account for a portion of the observed net elevation difference between localities.

A timberline 10 m below present at Cirque Peak between 1,400 and 950 yr BP may be a reflection of the shorter average lifespan of foxtail pine than bristlecone pine. This difference in longevity makes foxtail pine less likely to survive unfavourable periods, and thus more sensitive (ring width) and responsive (elevation change) to climatic variation than longer-lived species. Thus, the Cirque Peak record for this period may indicate cold conditions and a related glacial advance not readily apparent in other upper timberline records.

Mid- to late-Holocene variations of Cirque Peak timberline mirror the timing of major changes in timberline elevation in the White Mountains³ and, with the exception of the rise to current levels at 950 yr BP, the direction of specific timberline changes at Campito Mountain (Fig. 3). A similar decreasing trend in timberline elevation similar to that at Cirque Peak has been observed in the Rocky Mountains⁶. This correspondence supports a view of regional climatic forcing of timberline elevation. An additional correspondence of timberline decline with glacial advance in the Cirque Peak region⁷ suggests that the chronology of the Cirque Peak timberline can provide a link to the evaluation of climatic conditions leading to glacial advance and retreat in the Sierra Nevada and possibly the western United States.

Long, continuous and precisely dated records of timberline variation are uncommon and are especially valuable in the Sierra Nevada where, due to lack of absolutely datable material, little radiometric dating of climatically induced events such as glaciation has been possible. The Cirque Peak timberline and associated annual ring-width variation record, in conjunction with pollen and relative-age-dated deposits, will provide additional information about the physical environment in the Sierra Nevada over the past 6,300 yr. The integration of this record with those from the White Mountains will help isolate the

principal mechanisms responsible for climatic fluctuations during the mid to late Holocene and allow the evaluation of climatic spatial variability.

I thank Rainer Berger and David McJunkin of UCLA for the radiocarbon analysis and Donald Graybill, Rainer Berger and Joan Drake for critical review.

Received 25 June; accepted 17 October 1986.

1. Griggs, R. F. *Ecology* **27**, 275-289 (1946).
2. LaMarche, V. C. Jr & Mooney, H. A. *Nature* **213**, 980-982 (1967).
3. LaMarche, V. C. Jr *Quat. Res.* **3**, 632-660 (1973).
4. Carrara, P. E., Mode, W. N., Rubin, M. & Robinson, S. W. *Quat. Res.* **21**, 42-55 (1984).
5. Petersen, K. L. & Mehringer, P. J. Jr *Arctic Alpine Res.* **8**, 275-288 (1976).
6. Kearney, M. S. & Luckman, B. H. *Science* **221**, 261-263 (1983).
7. Scuderi, L. A. thesis, Univ. California, Los Angeles (1984).
8. Baillie, M. G. L. & Pilcher, J. R. *Tree-Ring Bull.* **33**, 7-14 (1973).
9. Suess, H. E. in *Radiocarbon Dating* (eds Berger, R. & Suess, H. E.) 777-784 (Univ. California Press, Berkeley, 1979).
10. Curry, R. R. *Spec. Pap. geol. Soc. Am.* no. 123, 1-47 (1969).
11. LaMarche, V. C. Jr & Stockton, C. W. *Tree Ring Bull.* **34**, 21-45 (1974).
12. LaMarche, V. C. Jr, Graybill, D. A., Fritts, H. C. & Rose, M. R. *Science* **225**, 1019-1021 (1984).

Decrease in precipitation acidity resulting from decreased SO_4^{2-} concentration

Lars O. Hedin^{*†}, Gene E. Likens[†]
& F. Herbert Bormann[‡]

^{*}Department of Biology, Yale University, New Haven, Connecticut 06511, USA

[†]Institute of Ecosystem Studies, The New York Botanical Garden, Box AB, Millbrook, New York 12545, USA

[‡]School of Forestry and Environmental Studies, Yale University, New Haven, Connecticut 06511, USA

The effect of decreases in SO_2 emissions on precipitation acidity has received much attention¹⁻¹⁰, but there has been no direct quantification of how recent decreases in SO_2 emissions in the northeastern and midwestern United States have affected precipitation acidity at a local site. Yet such quantification is an important step in assessing the effectiveness of control measures for SO_2 emissions. It is thought that recent decreases in SO_4^{2-} concentration in precipitation at the Hubbard Brook Experimental Forest (HBEF), New Hampshire, USA, result from decreases in SO_2 emissions. The effect of the SO_4^{2-} decrease on precipitation acidity is obscured by long-term trends in other ions, which also influence acidity. Here we show that, given the observed trends in concentration of other ions, the H^+ concentration in 1983 would have been nearly two-thirds higher than the measured values if the SO_4^{2-} concentration had not decreased.

During the 20-year period 1964-1983, decreases in concentration and deposition of SO_4^{2-} (Fig. 1a) in precipitation at HBEF have been attributed to decreases in SO_2 emissions in the northeastern and midwestern USA (Environmental Protection Agency regions I, II and V)^{10,11}. Significant decreasing trends in SO_4^{2-} concentrations have also been shown at 5 out of 8 National Atmospheric Deposition Program monitoring sites in the northeastern and midwestern USA since their inception in 1978¹². There is also evidence of a close relationship between SO_2 emission and SO_4^{2-} deposition from the western USA⁷, and indications of a similar relationship in Europe¹³.

Air masses from the Atlantic Ocean contribute <3% to the total SO_4^{2-} deposition at HBEF¹⁴. Locally derived soil dust is a small component¹⁵ and unlikely to contribute significant amounts of SO_4^{2-} . Atmospheric sulphur in the northeastern USA derives almost entirely from anthropogenic SO_2 emissions: this is further evidence that the decreases in SO_4^{2-} (Fig. 1a) at HBEF are due to decreases in emissions of SO_2 (refs 8-10, 16-19).

Here we quantify the effect of these decreases in SO_4^{2-} on precipitation acidity, using precipitation chemistry from bulk

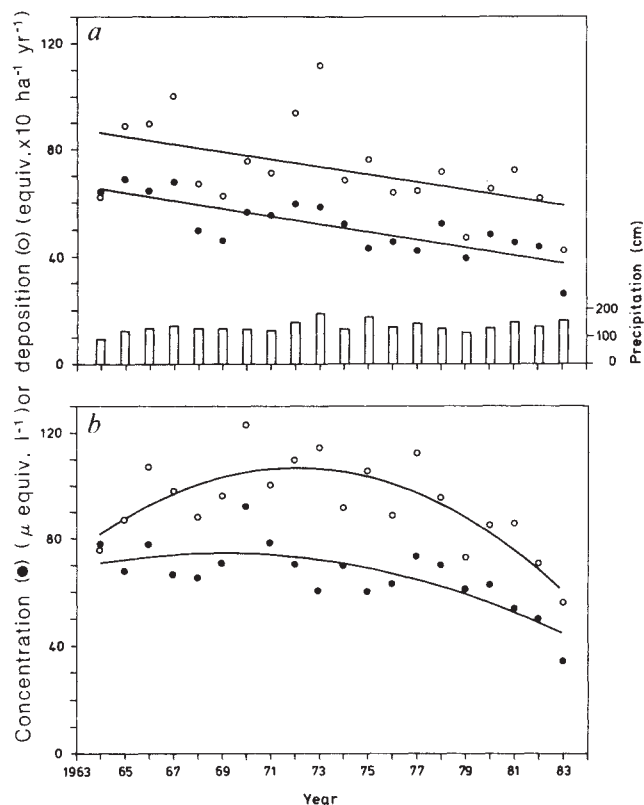


Fig. 1 Annual volume-weighted average ion concentration (●) and deposition (○) for SO_4^{2-} (a) and H^+ (b) in precipitation at watershed 6 of HBEF for 1964-83. Lines represent significant ($P \leq 0.05$) linear or curvilinear regression trends. The calculation of volume-weighted averages is described elsewhere²⁰. The histogram in a shows annual precipitation values at HBEF.

precipitation monitors during a 20-year period at HBEF. We emphasize, however, that although we assume that the decreases in SO_4^{2-} (Fig. 1a) at HBEF are a result of decreases in SO_2 emissions, our calculations are not dependent on an understanding of the specific shape of the relationship between SO_2 emission and SO_4^{2-} deposition, nor on a knowledge of the relative contribution of local versus distant emission sources. Our sampling and analytical methods have been described elsewhere²⁰.

Both concentration and deposition of SO_4^{2-} at HBEF show statistically significant ($P \leq 0.0001$ and $P \leq 0.03$) linear declines over the 20-year period (Fig. 1a, Table 1). The average rates of decrease in H^+ concentration and deposition, based on linear regression analysis of data from 1964 to 1983, are not statistically different from those of SO_4^{2-} ($0.9 > P > 0.5$; t -test). However, in contrast to SO_4^{2-} , the regression trends for H^+ concentration and deposition (Fig. 1b) have significant quadratic terms ($P \leq 0.03$ and $P \leq 0.0001$), and are best described by curvilinear regressions (Table 1). The trends in concentration and deposition of H^+ show modest increases during the beginning of the sampling period followed by sharp declines towards the end (Fig. 1b), whereas the SO_4^{2-} trends show constant rates of decrease throughout the period (Fig. 1a).

The absence of a precise correlation between long-term trends in SO_4^{2-} and in H^+ has led some authors to argue that the relationship between SO_4^{2-} and precipitation acidity is unclear, and that further reductions of SO_4^{2-} concentrations will not affect precipitation acidity^{5,6}. However, this ignores the complexity of the chemical relationship between precipitation acidity and SO_4^{2-} concentration. Decreases in SO_4^{2-} concentration would result in stoichiometric reductions in H^+ concentration only if other strong acid anions and base cations remain unchanged. In fact, precipitation acidity at HBEF has been

influenced by significant, and sometimes complex, trends in other strong acid anions (for example NO_3^-), and in basic materials available to neutralize strong acids (Table 1)^{21,22}. Thus we would not expect a stoichiometric correlation between the long-term trends of SO_4^{2-} and H^+ in precipitation at HBEF.

For example, although SO_4^{2-} concentration in precipitation decreased at HBEF during the first half of the sampling period (Fig. 1a), 1964 to 1975, the contribution of NO_3^- to precipitation acidity increased, and concentrations of base cations and Cl^- decreased. H^+ ion concentration remained fairly constant during this time (Fig. 1b), as potential H^+ decreases due to the SO_4^{2-} decline were balanced by trends in the other ions.

Because precipitation acidity at HBEF has also been affected by factors other than SO_4^{2-} trends, the relevant question is not whether trends in H^+ have mirrored decreases in SO_4^{2-} , but rather: 'What would the precipitation acidity at HBEF have been if SO_4^{2-} concentration had not decreased since 1964?'

To answer this question we calculated the acidity of precipitation at HBEF, assuming that SO_4^{2-} concentration did not decrease after 1964, and that all trends in other ions and the amount of precipitation remained as observed during the period 1964 to 1983. Predicted concentrations and deposition of H^+ for each year between 1964 and 1983 were calculated by holding SO_4^{2-} concentration constant at the 1964 regression value ($65.8 \mu\text{equiv. l}^{-1}$) and adding H^+ ions to the cation fraction to balance the increase in SO_4^{2-} . Each mole of SO_4^{2-} thus corresponds to two moles of H^+ , as is expected for the atmospheric conversion of SO_2 if other ions remain unchanged¹⁰. Other ions and precipitation volumes were kept at their original values for any specific year. Our calculation thus predicts the effect of SO_4^{2-} on precipitation acidity, taking into account the effect of long-term trends of other ions on precipitation acidity.

We found that if SO_4^{2-} concentrations in precipitation at HBEF had remained constant at the 1964 regression value, the H^+ concentration would have been $\sim 73 \mu\text{equiv. l}^{-1}$ in 1983 (Fig. 2a). But because there was an actual decrease in SO_4^{2-} concentration between 1964 and 1983, the measured H^+ concentration in precipitation at HBEF decreased significantly from a regression value of $71 \mu\text{equiv. l}^{-1}$ in 1964 to $45 \mu\text{equiv. l}^{-1}$ in 1983 (Fig. 2a). Thus without the SO_4^{2-} decrease, the average H^+ concentration in precipitation at HBEF in 1983 would have been nearly two-thirds higher than the actual measured concentration.

The variability in the predicted H^+ concentrations between different years (Fig. 2a) is due to the contribution of components other than SO_4^{2-} to the H^+ concentration, and to climatic variability. Although the predicted H^+ concentration shows consider-

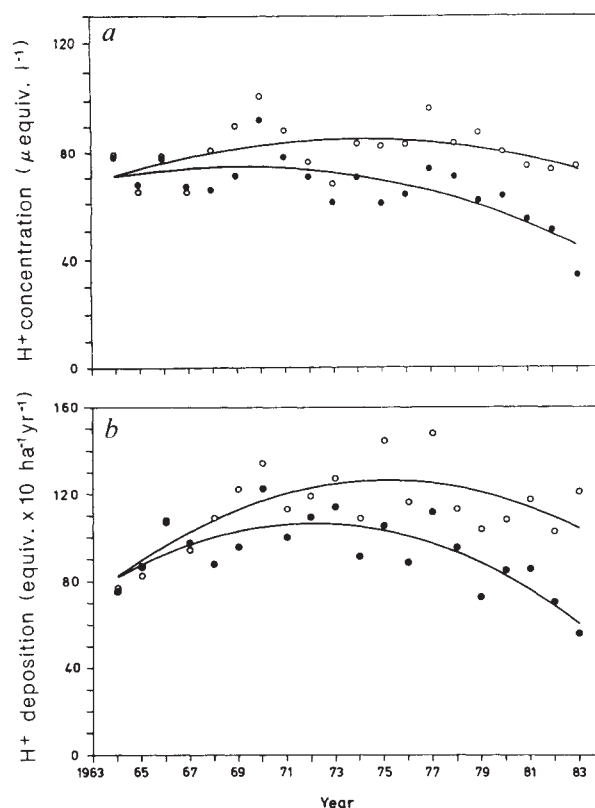


Fig. 2 Actual (●) and predicted (○) volume-weighted average H^+ concentration (a) and deposition (b) in precipitation at watershed no. 6 of HBEF during 1964–83. The predicted values assume no decrease in SO_4^{2-} concentrations since 1964. Lines represent regression trends. The predicted and actual trends are significantly different ($P \leq 0.01$; t -test).

able variability during certain periods of the record (for example 1970–73) and is best modelled by a slightly curved trend ($P \leq 0.05$), there is no significant linear decrease when the whole record from 1964 to 1983 is considered ($P > 0.7$).

Deposition of H^+ also would have been higher if SO_4^{2-} had not decreased after 1964. If SO_4^{2-} concentrations in precipitation at HBEF remained constant at the 1964 level, H^+ deposition would have increased from a regression value of $820 \text{ equiv. ha}^{-1} \text{ yr}^{-1}$ in 1964 to $1,040 \text{ equiv. ha}^{-1} \text{ yr}^{-1}$ in 1983 (Fig. 2b). But the

Table 1 Summary statistics of long-term trends in annual volume-weighted means of ions in precipitation at HBEF

Variable		Linear coefficient	s.e.	$P \leq$	Quadratic coefficient	s.e.	$P \leq$	R^2
SO_4^{2-}	Conc.	-1.49	0.25	0.0001	NS	—	—	0.664
NO_3^-	Conc.	0.0964	0.15	NS	-0.0920	0.028	0.005	0.389
Cl^-	Conc.	-0.723	0.23	0.008	NS	—	—	0.393
H^+	Conc.	-1.36	0.32	0.0004	-0.152	0.061	0.03	0.595
Base cations	Conc.	-1.20	0.20	0.0001	0.108	0.039	0.02	0.714
SO_4^{2-}	Depos.	-14.3	5.8	0.03	NS	—	—	0.255
NO_3^-	Depos.	3.12	2.5	NS	-1.68	0.48	0.003	0.442
Cl^-	Depos.	-9.98	3.2	0.008	NS	—	—	0.389
H^+	Depos.	-11.1	4.0	0.02	-3.84	0.78	0.0001	0.654
Base cations	Depos.	-12.4	2.3	0.0001	NS	—	—	0.625

Trends in SO_4^{2-} , NO_3^- , Cl^- , H^+ and base cations (sum of Na^+ , K^+ , NH_4^+ , Mg^{2+} , and Ca^{2+}) were modelled both as simple linear and as quadratic regressions. Linear regressions are of the form $Y = bX + a$, where b and a are coefficients, Y is ion concentration or deposition, and X is year. Quadratic regressions are of the form $Y = b_1X + b_2Z^2 + a$; where b_1 , b_2 and a are coefficients, Z^2 is a quadratic year term, corrected to orthogonality by subtraction of $\bar{X}$ ($Z = (X - \bar{X})$). Regression residuals were inspected. Only significant ($P \leq 0.05$) quadratic terms were included in the final model. The addition of an autoregressive term with a lag time of one year had little effect on standard errors and significance levels. PO_4^{3-} contributed less than 2% to total anion activity during 1964–83. The contribution of organic acids to free acidity in rainfall at HBEF is about 2%. Monitoring of Al^{3+} started in 1976. To avoid bias in the base cation fraction, Al^{3+} is therefore excluded from any of our trend interpretations. The average Al^{3+} volume-weighted concentration during 1976–83 was $1.8 \mu\text{equiv. l}^{-1}$.

s.e., standard error of coefficient; NS., not significant at $P \leq 0.05$ level; R^2 refers to model including significant terms only.

actual H^+ deposition at HBEF decreased significantly from a regression value of 820 equiv. $ha^{-1} yr^{-1}$ in 1964 to 610 equiv. $ha^{-1} yr^{-1}$ in 1983 (Fig. 2b). Without the decrease in SO_4^{2-} concentration, the average H^+ deposition at HBEF in 1983 would therefore have been about 70% higher than measured.

Our calculations assume that long-term trends in different base cations are not influenced by the long-term trend in SO_4^{2-} . This assumption is supported by the fact that base cations in precipitation come primarily from soil dust or from sea-spray rather than from anthropogenic emissions²³. However, if there were a link between base cation and sulphur chemistry, our results might require some revision. For example, trends of NH_4^+ in precipitation may be directly linked to trends in SO_4^{2-} (refs 23 and 24). But the ratio of NH_4^+ to SO_4^{2-} in precipitation at HBEF varies so widely between different years (max. 0.30; min. 0.13; C.V. = 21%) that this seems unlikely. Yet even if the decrease in NH_4^+ during 1964–83 in precipitation at HBEF were entirely dependent on the decrease in SO_4^{2-} , this would introduce a maximum bias of 5 $\mu\text{equiv. l}^{-1}$ to our calculations.

We conclude that the decrease in SO_4^{2-} content of precipitation at HBEF, generally attributed to decreases in SO_2 emissions, has resulted in large decreases in both concentration and deposition of H^+ , thereby diminishing the potential effects of acidification on organisms, water and soil.

Despite decreases, SO_4^{2-} concentrations in precipitation at HBEF remain 4–13 times higher than in remote areas receiving unpolluted rainfall²⁵. The regression value for SO_4^{2-} deposition in 1983 (28.5 $kg ha^{-1} yr^{-1}$) remains larger than the target loadings (14–16 $kg ha^{-1} yr^{-1}$) suggested as an upper limit for prevention of biological and chemical damage to aquatic ecosystems²⁴. Further decreases in SO_2 emissions are therefore likely to decrease SO_4^{2-} deposition and precipitation acidity at HBEF even more.

This study was supported by grants from the NSF, the Andrew W. Mellon Foundation and the Mary Flagler Cary Charitable Trust. This is a contribution to the Hubbard Brook Ecosystem Study, and to the Program of Institute of Ecosystem Studies. The HBEF is maintained by the USDA Forest Service. We thank J. Eaton, J. Hartigan and E. Gorham for consultation and B. Bormann, T. Butler, J. Cole, N. Caraco, G. Lovett, S. Odén, E. Gorham, J. Parker and K. Weathers for comments.

Received 30 May; accepted 16 October 1986.

1. Ableson, P. H. *Science* **230**, 617 (1985).
2. Ableson, P. H. *Science* **226**, 1263 (1984).
3. Oppenheimer, M. *et al. Science* **227**, 1155–1156 (1985).
4. Hidy, G. M. *et al. APCA JI* **31**, 333–354 (1984).
5. Katzenstein, A. W. *Science* **228**, 390–391 (1985).
6. Katzenstein, A. W. *Wall Street J.*, 28 June 1984.
7. Oppenheimer, M., Epstein, C. B. & Yunkhe, R. E. *Science* **229**, 859–862 (1985).
8. Ad Hoc Committee on Acid Rain: Science and Policy *Is there Scientific Consensus on Acid Rain?* (Institute of Ecosystem Studies, Millbrook, New York, 1985).
9. Lewis, D. & Davis, W. *Joint Report of the Special Envoys on Acid Rain*, Report to the President, 8 Jan. 1986.
10. National Research Council *Acid Deposition: Atmospheric Processes in Eastern North America* (National Academy Press, Washington, DC, 1983).
11. Gschwendtner, G., Gschwendtner, K. C. & Eldridge, K. *Historic Emissions of Sulfur and Nitrogen Oxides in the United States from 1900 to 1980*, USEPA Project Summary EPA/600/S7-85/009 (1985).
12. Schertz, T. L. & Hirsch, R. M. *Trend Analysis of Weekly Acid Rain Data—1978–83*, USGS Water Resources Investigation Report 85–4211 (1986).
13. Rodhe, H. & Granat, L. *Atmos. Environ.* **18**, 2627–2639 (1984).
14. Munn, R. E. *et al. A Climatological Analysis of the Hubbard Brook (New Hampshire) Precipitation Chemistry Data* (Institute for Environmental Studies, Univ. Toronto, 1982).
15. Likens, G. E. *et al. Ecology* **48**, 772–785 (1967).
16. *The Acidic Deposition Phenomenon and Its Effects: Critical Assessment Review Papers*, Vols I and II, USEPA Report EPA-600/8-83-016A (1984).
17. National Research Council *Atmospheric-Biosphere Interactions: Toward a Better Understanding of the Consequences of Fossil Fuel Combustion* (National Academy Press, Washington DC, 1981).
18. *Report of the Acid Rain Peer Review Panel* (Office of Science and Technology Policy, Washington DC, 1984).
19. *Acid Rain and Transported Air Pollutants: Implications for Public Policy*, Report OTA-O-204 (Office of Technology Assessment, US Congress, Washington DC, 1984).
20. Likens, G. E. *et al. Biogeochemistry of a Forested Ecosystem* (Springer, New York, 1977).
21. Likens, G. E. *et al. Atmos. Environ.* **18**, 2641–2647 (1984).
22. Galloway, J. N. & Likens, G. E. *Atmos. Environ.* **15**, 1081 (1981).
23. Munger, J. W. & Eisenreich, S. J. *Environ. Sci. Technol.* **17**, 32A–42A (1983).
24. Gorham, E. *et al. Science* **225**, 407–409 (1984).
25. Galloway, J. N., Likens, G. E. & Hawley, M. E. *Science* **226**, 829–831 (1984).

Free amino acids in marine rains: evidence for oxidation and potential role in nitrogen cycling

Kenneth Mopper & Rod G. Zika

University of Miami, Rosenstiel School of Marine and Atmospheric Science, Division of Marine and Atmospheric Chemistry, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA

Previous studies of dissolved organic nitrogen (DON) in precipitation have addressed various aspects of nutrient transport and global nitrogen cycling¹. In most of these studies however, the detailed chemical composition of DON was not determined. Analyses of specific organic nitrogen compounds within precipitation can yield new information about sources and transformations of DON as well as about heterogeneous oxidative processes in the atmosphere. Dissolved free amino acids (DFAA) are a class of compounds for which surprisingly few analyses in precipitation have been published¹. We report here the first detailed analyses of DFAA and primary amines in marine rains. Unexpectedly high concentrations of total DFAA were measured, averaging about 6.5 μM (16 marine rain samples) and ranging from 1.1 to 15.2 μM (0.015 to 0.21 p.p.m. N); these values are similar to those obtained for inorganic nitrogen species². Amino acids are predominantly found in rain as their L optical isomers and therefore are most likely biological in origin; however, the exact sources and modes of enrichment in rain over the open ocean are not known. There is evidence that some of the amino acids, in particular methionine, are oxidized in the atmosphere possibly by a heterogeneous photochemical pathway. The finding of high DFAA and primary amine concentrations in marine rain may have important implications in global nitrogen cycling and also may contribute locally to available nitrogen at the sea surface.

Rain, air, and seawater samples were analysed at sea for one year during four recent cruises and at the home laboratory in Miami as part of the SOLARS programme (Studies of Light Activated Reactions in the Sea). These cruises were located in the Gulf of Mexico (RV *Cape Florida*) and in different locations in the northwest Atlantic Ocean (RV *Columbus Iselin*) (Table 1). Amino acids, primary amines and ammonium were analysed on board by high performance liquid chromatography in conjunction with precolumn fluorescence derivatization with o-phthalaldehyde and a thiol³. For some samples two different thiols, 2-mercaptoethanol (ME) and N-acetyl-L-cysteine (NAC), were employed in separate derivatizations. This was done to verify the identification of specific amino acids, since the thiol used in the reaction greatly alters the chromatographic selectivity of the fluorescent derivatives. In addition, the use of NAC results in the separation of L and D enantiomers⁴ and also permits more sensitive detection of ammonium than ME. Secondary and tertiary amines cannot be determined by this technique.

Extreme care was exercised to eliminate contamination during sampling. Rain samples were collected with acid washed glass funnels and bottles that were rinsed several times with rain before the actual samples were taken. Air was sampled using an all glass impinger filled with deionized water; the air flow was 1–2 litre min^{-1} and approximately 0.1–0.2 m^3 of unfiltered air were sampled. Most rain samples were analysed unfiltered and within one hour of collection. Several samples were filtered (0.7 μm glass fibre or 0.2 μm membrane) and stored frozen. Filtration and freezing were found not to alter significantly the composition of samples. A summary of the results and background information on a few representative rain and air samples are given in Table 1; detailed DFAA and primary amine compositions of these samples are given in Table 2. Concentrations up to 15 μM were observed in marine rains. Procedural blanks (deionized water or sea water put through the above collection

Table 1 Amino acids, primary amines, ammonium and background data for rain and air samples

Sample	Date	Location	Local Time (EST)	Prevailing wind direction	Amino acids (μM)	Primary amines (μM)	NH_4^+ (μM)	NO_3^- (μM)	Cl^- (μM)	SO_4^{2-} (μM)	H_2O_2^* (μM)
<i>Gulf of Mexico</i>											
B	9/16/85	26°22' N 89°52' W	19:30	NNE	13.22	0.10	13.24	18	3,024	191	38.6
D	9/19/85	24°38' N 82°57' W	12:17	E	15.17	0.08	ND	ND	ND	ND	20.2
<i>Miami, Florida</i>											
F	10/2/85	Miami	11:00	SE	0.61	0.25	1.2	282	3,945	366	ND
I-3			17:10	NW	0.39	0.035	3.79	30	82	12	ND
I-3 Hydrolysis			17:10		3.03	1.32	13.14	ND	ND	ND	ND
<i>Northwestern Atlantic Ocean</i>											
O-3	2/23/86	25°55' N 78°13' W	22:55–23:21	SW	10.80	1.36	13.06	1,169	666	52	8.4
Q-1	2/26/86	20°31' N 68°44' W	10:25–10:30	N	4.91	0.22	16.36	80	883	70	18.7
T	6/22/86	26°46' N 75°24' W	05:45–06:00	E	1.71	0.12	2.81	23	206	ND	ND
U	9/26/86	26°30' N 76°04' W	19:00–19:05	E	4.04	0.65	17.52	ND	ND	ND	80
Marine air 1†	2/28/86	13°12' N 66°04' W	11:45–16:40	E	2.18	0.022	1.08	527	265	29	ND
Surface sea water (μM)	10/25/86	26°30' N 76°04' W	14:40	E	0.062	—	0.42	ND	ND	ND	0.012

ND, not determined.

* See ref. 17.

† Concentrations in nmol m^{-3} .

and sample handling procedures) showed no evidence of amino acid contamination from either the sampling personnel or the ship. Furthermore, the sample handling and derivatization procedures were identical to those routinely used on the same cruises for the determination of DFAA and ammonium in seawater, where total concentrations were 2–3 orders of magnitude lower than in most marine rain samples (Table 2).

The DFAA pattern of marine rain is dominated by glycine, serine, alanine and the acidic amino acids (Table 2). Close examination of the spectra reveals the presence or absence of certain amino compounds that suggest that DFAA are being abiotically degraded in the atmosphere. For example, the absence of methionine in many samples is especially noteworthy, as it is a common constituent of proteins in all organisms. Just as striking is the presence of methionine sulfoxide in nearly all rain and air samples because this amino acid has very few biological sources. Furthermore, methionine sulfoxide eluted as two peaks corresponding to L-methionine-L-sulfoxide and L-methionine-D-sulfoxide. The above results may be explained in terms of abiotic oxidation of methionine, probably within aerosols, resulting in a racemic mixture of the sulfoxides. Previous studies have shown that methionine can be oxidized by several pathways^{5,6} and methionine sulfoxide is a major product⁶. We are currently conducting experiments under environmentally relevant conditions to study potential pathways and oxidants involved. First results suggest that reactive species produced during the photodecomposition of potential atmospheric sources of oxidants, such as nitrate and hydrogen peroxide⁷, readily oxidize the reduced sulphur in methionine. The high concentrations of primary amines, especially 2-aminoethanol and methylamine, in several rain samples are also noteworthy (Table 2). Experiments are being planned to test whether these species can also be derived from abiotic alteration (decarboxylation, for example) of amino acids under the prevailing conditions in marine aerosols. Details of these and related experiments will be presented elsewhere.

With the exception of methionine sulfoxide, only L amino acids were detected, indicating a biological origin for these compounds in rain as opposed to abiotic syntheses as suggested

by other investigators^{8,9}. Although biological in origin, the exact sources, and modes of transport and enrichment of DFAA and primary amines in marine rains are not known at present.

It is evident from Table 1 that marine rains are strongly enriched in DFAA (on average by a factor of ten) relative to terrestrial rains. In fact, it is likely that this enrichment is even more pronounced than our results suggest because rains in Miami also have a marine component. In support of this, Likens *et al.*¹⁰ reported that the mean concentration of total dissolved amines in precipitation at two inland stations was $0.060 \mu\text{M}$ (about ten times lower than rains collected in Miami). Hydrolysis

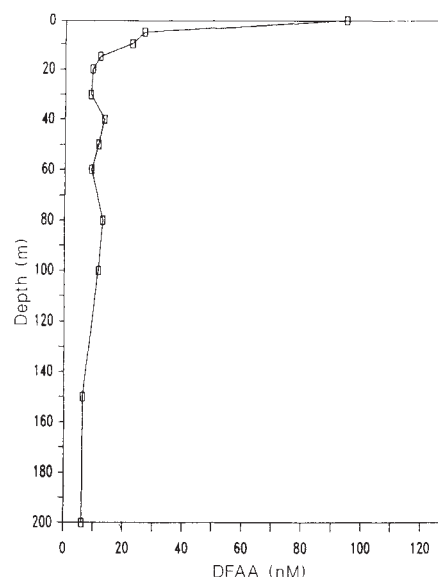


Fig. 1 Effect of rain on the concentration of dissolved free amino acids in the water column at a station in the Gulf of Mexico (27° 10' N, 89° 02' W; 9/14/85; RV *Cape Florida* cruise CF8512; station 4). The analyses were performed on board using HPLC, as described in the text. The total DFAA concentration in the rain was approximately $4.5 \mu\text{M}$.

Table 2 Free amino acids, primary amines and ammonium in marine rain and related samples (mole % composition and total concentration)

Compound†	Sample* (mole % composition)										Surface seawater (NAC)
	B‡ (NAC)	D (ME)	F (NAC)	I-3 (NAC)	I-3 (hydrolysed)	O-3 (NAC)	Q-1 (ME)	T (NAC)	U (NAC)	Marine air 1	
Amino acid											
CysA	—	—	ND	ND	ND	ND	ND	ND	ND	ND	ND
Asp	6.2	4.9	0.7	6.7	10.7	5.9	7.7	10.0	3.4	8.3	9.8
Glu	4.6	2.4	2.6	13.4	17.2	3.0	6.6	—	3.8	1.8	7.1
BAGA	tr	—	—	—	—	0.1	—	—	—	—	—
Asn	0.7	0.7	0.5	2.1	—	0.6	0.6	0.6	1.2	1.1	3.2
Ser	24.1	26.4	31.8	13.4	9.4	15.8	15.9	22.3	24.5	26.4	22.4
Gln	0.8	0.2	0.3	2.1	0.4	0.9	0.5	0.6	0.7	1.6	1.6
His	ND	3.9	2.1	ND	—	ND	ND	ND	ND	ND	ND
Thr	5.6	1.8	7.6	4.9	3.2	5.6	2.4	6.2	5.2	3.6	4.0
Gly	18.8	24.4	21.8	15.9	35.7	27.2	17.4	20.0	20.5	22.3	22.6
MetSO	1.8	0.2	—	0.8	0.5	0.7	1.1	0.4	0.3	0.9	—
BAla	—	tr	ND	—	0.5	tr	tr	—	—	0.2	—
Cit	2.9	2.2	0.5	1.8	—	1.1	0.8	2.9	2.5	0.6	1.5
Tau	0.4	0.2	0.5	1.5	0.6	0.3	0.4	0.4	1.7	0.4	1.6
Arg	3.0	0.9	0.7	3.9	2.6	0.9	1.0	1.4	—	0.9	—
GABA	0.5	ND	0.3	2.1	0.7	0.2	8.4	0.2	0.8	1.1	—
MeHis	tr	ND	—	—	0.1	—	—	—	—	—	—
Ala	8.4	15.2	10.6	10.5	6.0	13.4	10.2	9.4	8.2	10.6	7.3
BABA	—	—	—	—	—	—	—	—	—	tr	—
Tyr	3.1	2.1	ND	2.3	1.1	3.0	1.5	2.2	2.0	0.6	2.9
AABA	—	—	—	—	—	—	—	—	—	1.3	—
Val	3.9	3.4	5.6	4.1	2.5	6.3	2.4	3.4	4.0	3.2	4.5
Met	—	—	—	—	0.1	0.2	0.9	—	0.2	0.7	—
Trp	2.0	0.1	—	0.8	—	0.2	0.3	0.2	0.1	tr	—
Hyl	—	ND	—	—	—	0.7	0.6	0.3	0.4	1.2	—
Phe	1.9	1.3	1.5	2.3	1.5	2.8	1.4	1.6	2.7	2.6	1.3
Ile	2.5	1.9	2.0	2.3	1.6	3.4	1.6	2.6	2.4	2.3	2.7
Orn	3.9	5.0	4.0	3.1	1.0	8.1	4.4	9.4	7.2	3.9	4.0
Leu	4.0	2.0	3.0	4.1	2.6	4.0	2.0	2.2	3.1	3.0	1.6
Lys	2.3	1.4	4.0	2.1	1.8	3.0	2.1	3.5	5.0	1.4	2.3
Total AA (μM)	13.80	15.17	0.61	0.39	3.03	10.80	4.91	1.71	4.04	2.18§	0.062
2AE	53.0	67.5	57.8	48.6	96.0	93.9	62.7	52.5	69.0	31.8	82.5
MA	47.0	28.9	40.6	40.0	2.0	4.7	28.1	47.5	25.3	68.2	17.5
EA	ND	3.6	1.6	11.4	2.0	1.4	9.2	NA	5.7	tr	—
Total primary amine (μM)	0.15	0.08	0.25	0.035	1.32	1.36	0.22	0.12	0.65	0.022§	0.004
NH ₄ ⁺ (μM)	13.24	ND	1.2	3.79	13.14	13.06	16.36	2.8	17.52	1.08§	0.42

* Sample descriptions for A, B, ... are given in Table 1; ME, 2-mercaptoethanol derivative; NAC, *N*-acetylcysteine derivative.

† Non-standard abbreviations: CysA, Cysteic acid; BAGA, β -aminoglutaric acid; MetSO, methionine sulphoxide; BAla, β -alanine; MeHis, methyl-histidine; BABA, β -aminobutyric acid; AABA, α -aminobutyric acid; Hyl, hydroxylysine; 2AE, 2-aminoethanol; MA, methylamine; EA, ethylamine.

‡ Sample stored at -20°C .

§ Concentration unit is nmol m^{-3} air.

|| Asp + Glu coeluted in sample T; Tau + Arg coeluted in sample U.

of unfiltered terrestrial rains (as in Table 2 sample I-3), reveals that the concentrations of combined amino acids and primary amines (principally 2-aminoethanol) are respectively tenfold and fortyfold higher than those of free constituents. Hydrolysis of an unfiltered marine rain on the other hand showed only about twofold and fourfold excesses of combined amino acids and primary amines, respectively, over free constituents. A possible explanation for these results is that land-derived proteinaceous materials (for example, plant fragments, bacteria, yeasts, spores, pollen) associated with dust become leached and hydrolysed within the harsh chemical environment of aerosols as they are transported over the oceans. Expected products would be DFAA, primary amines and ammonium, which then become scavenged by rain. A terrestrial source would be consistent with the findings of other investigators who have shown that other organic constituents (for example, waxes and other lipids) in marine aerosols are in part derived from land sources¹¹.

Alternatively, the ocean itself may be the main source for the observed enrichment. High concentrations, up to $14\text{ }\mu\text{M}$ protein-N, were reported in marine rains collected in New Zealand and were speculated to be derived from proteinaceous substances in sea spray¹². In fact, aerosols experimentally generated from

seawater have been shown to be enriched by 100–1,000 fold in organic carbon and organic nitrogen relative to bulk seawater^{13–15}. These aerosols contained 0.5–2.7 mM protein-N¹³, while sea foam contained up to 16.8 mM protein-N¹². From these past studies, a marine origin for the high concentrations of DFAA in our marine rain samples seems highly probable. However, more sampling over a broader geographical area and measurements of parameters such as stable isotopes and DFAA in size-fractionated aerosols must be accomplished before the relative contribution of marine versus terrestrial sources can be estimated.

If the ocean is the major source for the high concentrations of DFAA and primary amines in marine rains, our results suggest that these and possibly other organic nitrogen compounds are largely recycled over the ocean, although they may be transported somewhat inland¹². On the other hand, if nitrogenous substances are derived mainly from land sources, then the results imply that there is a net transport of new, biologically utilizable organic nitrogen to the oceans. In either case, the concentrations of DFAA in marine rains are sufficiently high that they can locally augment the concentration of available nitrogen at the sea surface and perhaps induce a local enhancement of produc-

tivity, as recently suggested for inorganic nitrogen from rain¹⁶. Analyses of the water column during a rain event clearly reveal a surface enrichment, Fig. 1. It is unlikely that this enrichment was due mainly to *in situ* biological activity since the upper water column appeared to be well mixed at the time of sampling and the primary productivity maximum was not at the surface but at about 30–50 m in the water column (inferred from the chlorophyll *a* maximum). Furthermore, during the same cruises rain related enrichments at the sea surface were also observed for hydrogen peroxide¹⁷, which, like DFAA, is strongly enriched in marine rain relative to seawater (Table 1).

The finding of high concentrations of DFAA in rain suggests that high molecular weight organic nitrogen (presumably proteinaceous) substances are degraded by chemical processes in the atmosphere (probably within aerosols) to low molecular weight species, such as DFAA, primary amines and ammonium. Evidence exists that some of these species are further degraded (oxidation of methionine). Therefore, assuming that DFAA and other organic nitrogen compounds are widely distributed in the marine troposphere at levels similar to those reported here, it is possible that these species may play an important role in the chemistry of atmospheric inorganic nitrogen especially if they can be oxidized all the way to NO_3^- through as yet unknown atmospheric oxidative pathways. If this is possible, it would explain in part the relatively uniform background of aerosol

NO_3^- over large areas of the tropics¹⁸.

We thank W. J. Cooper, D. J. Kieber, C. Lee, J. M. Prospero and E. S. Saltzman for valuable discussions and M. P. Daniel, R. Nees, C. I. Ross, and R. J. Sikorski for technical assistance. We also thank Captain Morgan and the crews of the RV *Cape Florida* and *Columbus Iselin* for their expert help during sampling. We gratefully acknowledge the NSF (OCE-8411781 and ATM-8405291) and the Office of Naval Research (N00014-85-C-0020) for providing shiptime and other support for this study.

Received 4 August; accepted 13 November 1986.

1. Mazurek, M. A. & Simoneit, B. R. T. *Crit. Rev. Environ. Control* **16**, 1–140 (1986).
2. Knap, A., Jickells, T., Pszeny, A. & Galloway, J. *Nature* **320**, 158–160 (1986).
3. Mopper, K. & Lindroth, P. *Limnol. Oceanogr.* **27**, 336–347 (1982).
4. Nimura, N. & Kinoshita, T. *J. Chromatogr.* **352**, 169–177 (1986).
5. Nilsson, R., Merkel, P. B. & Kearns, D. R. *Photochem. Photobiol.* **16**, 117–124 (1972).
6. Sysak, P. K., Foote, C. S. & Ching, T.-Y. *Photochem. Photobiol.* **26**, 19–27 (1977).
7. Graedel, T. E. & Weschler, C. J. *Rev. Geophys. Space Phys.* **19**, 505–539 (1981).
8. Fonselius, S. *Tellus* **6**, 90 (1954).
9. Sidle, A. B. *Tellus* **19**, 128–135 (1967).
10. Likens, G. E., Edgerton, E. S. & Galloway, J. N. *Tellus* **35B**, 16–24 (1983).
11. Duce, R. A. *et al. Rev. Geophys. Space Phys.* **21**, 921–952 (1983).
12. Wilson, A. T. *Nature* **184**, 99–101 (1959).
13. Gershey, R. M. *Limnol. Oceanogr.* **28**, 309–319 (1983).
14. Hoffman, E. J. & Duce, R. A. *J. geophys. Res.* **79**, 4474–4477 (1974).
15. Hoffman, E. J. & Duce, R. A. *J. geophys. Res.* **81**, 3667–3670 (1976).
16. Pearl, H. W. *Nature* **315**, 747–749 (1985).
17. Cooper, W. J., Saltzman, E. S. & Zika, R. G. *J. geophys. Res.* (in the press).
18. Prospero, J. M., Savoie, D., Nees, R. T., Duce, R. A. & Merrill, J. J. *J. geophys. Res.* **90**, 10,586–10,596 (1985).

Left hemisphere advantage in the mouse brain for recognizing ultrasonic communication calls

G. Ehret

Fakultät für Biologie, Universität Konstanz, Postfach 5560,
D-7750 Konstanz, FRG

In humans, sound perceived as speech is processed preferentially by the right ear and the left hemisphere of the brain^{1–3}. Among animals, such an advantage of one hemisphere (lateralization) in processing communication sound from other members of the species has so far been demonstrated only in macaque monkeys^{4–6}. I report here that in the house mouse, which has a very much less elaborate forebrain than man or macaque monkey, the ultrasonic calls that are emitted by young mice to evoke maternal caring behaviour are preferentially recognized by the left hemisphere. In females with no experience of pups, which have been trained to respond to the same ultrasonic calls by conditioning, no advantage for one hemisphere is detected. The results suggest that lateralization of this function evolved early in mammals and emphasize that an innate predisposition for perceiving communication sounds is connected with a left-hemisphere advantage in processing them. This experimental system is a readily-available animal model for studying lateralized auditory brain functions.

Young mouse offspring (pups) in discomfort emit ultrasonic calls (Fig. 1A) which are reliable releasers of maternal behaviour in their mothers, for example a calling pup displaced from the litter is retrieved to the nest⁷. A series of sound bursts with energy concentrated in a narrow frequency band (maximum bandwidth 22.5 kHz at a central frequency of 50 kHz) in the ultrasonic range is both necessary and sufficient for the release of maternal behaviour (Fig. 1A)⁸. Such ultrasounds are perceived categorically in the frequency domain^{9,10}, so that ultrasounds from a continuum of varying bandwidths are labelled as belonging to different classes, and sounds labelled differently are well discriminated whereas others labelled the same are less well discriminated. The attachment of the correct category labels (preferred or nonpreferred releaser) to a perceived sound has to be learned¹¹. This means that mice have an innate ability to

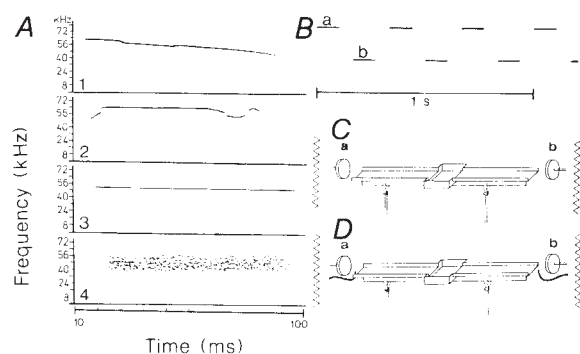


Fig. 1 A, Sonagram examples of two natural pup ultrasounds (1, 2) and of the two synthesized stimuli used as artificial releasers: 50 kHz tone bursts (3) and 40–60 kHz noise bursts (4). The artificial stimuli were of 80 ms duration (plus 10 ms rise and fall times), had a repetition rate of 3 s^{-1} and an intensity of 57 dB loudness level in the nest depression (equals 75 dB SPL re. $20 \mu\text{Pa}$ for 50 kHz tone and 40–60 kHz noise bursts and 61 dB SPL for 20 kHz tone bursts). B, Diagram showing the alternating series of sound stimuli from loudspeaker a and b. Bar, 1 second. C, Position of running board and speakers a and b in the first experiment. D, Same as in C but with water spouts at the end of the running board for giving the water reward in the second experiment.

categorize pup ultrasounds but have to learn the communicative significance ('meaning') of each category through experience with the pups.

Lateralization of ultrasound perception in the mouse brain was tested behaviourally in a two-alternative choice test⁸. Briefly, primiparous lactating house mice (outbred strain NMRI) with their litters (2–6 days old) were placed in the central nest depression of a running board (Fig. 1C). This was located in a sound-proof room under dim red light and the mice allowed to adapt to the conditions for at least six hours. Immediately before a test of ultrasound perception, between four and six pups were taken from the litter and distributed on the running board, so that maternal motivation was raised to a comparably high level in all females by the process of retrieving the pups. After the pups had been retrieved, the two ultrasonic loudspeakers (a and b in Fig. 1C) were switched on simultaneously (Fig. 1B), one emitting signals resembling natural releasers of maternal

Table 1 Responses of naive females in the conditioned discrimination of 50 versus 20 kHz tone bursts

Day	Test	20	25	28	Animal no. 29	30	34	36	38	Summed responses
1	Binaural (3 days)	19:1	17:3	16:4	13:7	14:6	13:7	15:5	15:5	122:38
2		17:3	16:3	16:4	14:6	14:6	15:5	13:7	16:4	121:39
3		18:2	18:2	15:5	15:5	13:7	13:7	14:6	14:6	120:40
4, 6	Right ear plugged	16:4*	13:7†	16:4†	13:7†	10:10*	14:6*	15:5†	11:9*	108:52
5	Binaural	14:6	15:5	16:4	17:3	14:6	13:7	18:2	15:5	122:38
4, 6	Left ear plugged	13:7†	15:5*	14:6*	14:6*	13:7†	14:6†	13:7*	14:6†	110:50
7	Binaural	13:7	16:4	17:3	15:5	13:7	17:3	15:5	16:4	122:38

Responses in 7 consecutive days of testing are shown as (number of responses to 50 Hz tone bursts):(number of responses to 20 kHz tone bursts).

* Tested on day 4.

† Tested on day 6.

behaviour (the artificial pup call stimulus: 50 kHz tone bursts or 40–60 kHz noise bursts), the other emitting neutral signals (20 kHz tone bursts not resembling pup calls)^{8,9}. In the discrimination tests, females were assessed as preferring one of these stimuli if they left the central nest area and moved at least 30 cm towards one of the speakers. Every female was tested a maximum of six times, with stimuli assigned to the speakers at random.

First, stimulus preference was tested binaurally. Then half the 22 females in each group (with either the 50 kHz tone or 40–60 kHz noise used as the artificial pup call) had temporary plugs put into their right ears, half had plugs put in their left ears, and all were re-tested. Immediately after the tests, the plugs were switched to the other ear. Final tests were run the next day. The earplugs (Ohropax packed into the ear canal and the pinna fixed with Loctite glue over the ear canal entrance under ketamine anaesthesia) attenuated the applied sound stimuli by at least 40 dB as measured by evoked potentials from the inferior colliculus.

Figure 2A shows the percentages of runs towards the 50 kHz tone bursts and the 40–60 kHz noise bursts. These ultrasounds are clearly preferred stimuli ($P < 0.01$ in a two-sided binomial test) in the binaural condition and with the left ear plugged. With the right ear closed, however, no preference is seen and runs are distributed at near chance levels. This means that the females with a closed right ear do not discriminate between the ultrasounds resembling pup calls and the 20 kHz control stimuli. This remarkable loss of preference can be explained in two ways: either the females are unable to discriminate the locations of the speakers with only the left ear functioning, or the mice are unable to determine the nature of the sound signals and the ultrasounds resembling pup calls can no longer be identified as significant releasers of maternal behaviour.

A second experiment was carried out to test these explanations and to examine the hypothesis that there is a right ear advantage only for sounds important in communication. Virgin females of the same age as the mothers (3–5 months) without pup experience, which do not identify ultrasounds as relevant releasers of maternal behaviour¹¹, were trained to prefer 50 kHz to 20 kHz tone bursts using water as a reward. Test conditions were identical with those in the first experiment except that no pups were involved and water spouts were placed at both ends of the running board to reward correct responses (Fig. 1D). Eight females learned to prefer the 50 kHz stimulus after an average of 18 days conditioning. When they had reached a criterion of at least 65% responses to 50 kHz in 20 runs per day on three consecutive days, the animals were tested with one ear plugged. Table 1 shows the results of the choices (50 kHz versus 20 kHz) of all individuals over seven days of testing. All but two animals preferred the 50 kHz tone bursts in at least 65% of the responses under all conditions of binaural and monaural testing, and thus did not show an ear advantage in this test of conditioned discrimination. In Fig. 2B the percentages of runs towards the 50 kHz stimulus, calculated from the summed responses of all animals, are shown separately for the seven test

days. Tone bursts of 50 kHz frequency are clearly preferred ($P < 0.0001$ in a two-sided binomial test), whether the animals listened with left, right or both ears.

This second experiment demonstrates that female house mice listening with either the right or left ear can locate a loudspeaker, discriminate between two sound stimuli of differing significance and connect the rewarded stimulus with a speaker at the appropriate location.

These two experiments therefore reveal a fundamental difference in sound perception. Mothers with pups do not identify the ultrasound stimuli as pup calls when they can use only their left ears, but they can do with their right alone. This right-ear (left-hemisphere) advantage does not occur when females with no experience of pups are conditioned to respond preferentially to the same ultrasounds (50 kHz tone bursts). Both experiments were conducted in the same experimental setting, with the same behavioural task and with the same stimuli (50 kHz and 20 kHz tone bursts); in both the 50 kHz stimulus indicated reward (pups could potentially be retrieved or a water reward could be obtained). Differences in response presumably occurred because the mothers had learned to identify the test stimuli through experience with their pups, and responded in

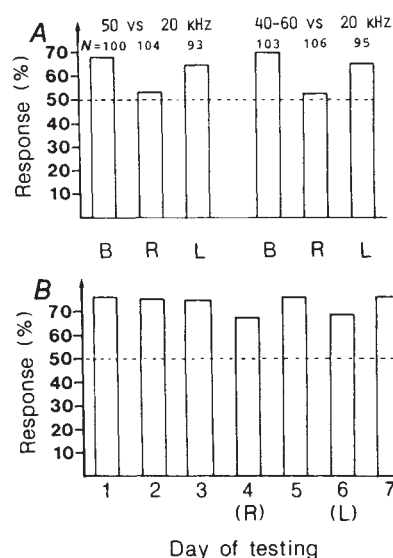


Fig. 2 A, Percentage of responses of lactating females to 50 kHz tone and 40–60 kHz noise bursts in the choice tests against 20 kHz tone bursts. Ears tested indicated thus: B, binaural; R, right ear plugged; L, left ear plugged. N, numbers of tests in each column. B, Percentage of responses of naive females to 50 kHz tone bursts in the choice tests against 20 kHz tone bursts. Tests were binaural, except R, right ear plugged and L, left ear plugged, and days 4 and 6 show data from all animals of the indicated test conditions although half of the animals were tested with the right ear plugged on day 4, the other half on day 6 (left ear vice versa).

a communicative context in the tests, whereas the naive females learned and responded in a plain instrumental situation not involving intraspecific communication.

Thus, an innate releasing mechanism activated in a communicative context preferentially in the left hemisphere of the brain is the most probable explanation for lateralized recognition of ultrasound in house mice. This interpretation is consonant with those put forward for the left-hemisphere advantage for species-specific call recognition in monkeys⁴⁻⁶ and with general views of hemispheric laterality in man² and animals¹². It should be considered as a possible basis of the left-hemisphere advantage for speech sound recognition in man.

Supported by the Deutsche Forschungsgemeinschaft, Eh 53/8,9. I thank Drs R. B. Freeman and M. Plummer for comments on the manuscript, Dr F. Gruber and his staff for animal care and F. Fischer, W. Hemmer and E. Wintergerst for running most of the tests.

Received 31 July; accepted 24 October 1986.

1. Cutting, J. E. *Percept. Psychophys.* **16**, 601-612 (1974).
2. Bradshaw, J. L. & Nettleton, N. C. *Behav. Brain Sci.* **4**, 51-91 (1981).
3. Geschwind, N. & Galaburda, A. M. *Archs Neurol.* **42**, 428-459 (1985).
4. Petersen, M. R., Beecher, M. D., Zoloth, S. R., Moody, D. B. & Stebbins, W. C. *Science* **202**, 324-327 (1978).
5. Petersen, M. R. *et al. Behav. Neurosci.* **98**, 779-790 (1984).
6. Heffner, H. E. & Heffner, R. S. *Science* **226**, 75-76 (1984).
7. Haack, B., Markl, H. & Ehret, G. in *The Auditory Psychobiology of the Mouse* (ed. Willott, J. F.) 57-97 (Thomas, Springfield, 1983).
8. Ehret, G. & Haack, B. *J. comp. Physiol.* **148**, 245-251 (1982).
9. Ehret, G. & Haack, B. *Naturwissenschaften* **68**, 208 (1981).
10. Ehret, G. in *Categorical Perception* (ed. Harnad, S.) (Cambridge University Press, Cambridge, in the press).
11. Koch, M. thesis, Univ. Konstanz (1986).
12. Denenberg, V. H. *Behav. Brain Sci.* **4**, 1-49 (1981).

A steroid sex pheromone synchronizes male-female spawning readiness in goldfish

J. G. Dulka, N. E. Stacey, P. W. Sorensen & G. J. Van Der Kraak

Department of Zoology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Understanding of pheromone function in teleost fish has been impeded by a lack of information on pheromone identities¹. Our recent studies^{2,3} on goldfish *Carassius auratus*, however, provide strong evidence that $17\alpha,20\beta$ -dihydroxy-4-pregnen-3-one (17,20P), the proposed oocyte maturation-inducing steroid hormone in goldfish⁴ and other teleosts⁵, could be a potent female sex pheromone. Milt (sperm and seminal fluid) volume in goldfish is increased by exposure to 17,20P (and to a lesser extent by exposure to two precursors of 17,20P, progesterone and 17α -hydroxyprogesterone) but not to other steroids proposed as fish pheromones^{1,6,7}. In addition, the goldfish olfactory epithelium is extremely sensitive to 17,20P (ref. 3), and the increase in milt volume normally induced by 17,20P exposure is abolished by sectioning the medial olfactory tracts, which previously have been implicated in the control of sex behaviour in male goldfish⁸. We report here that ovulating goldfish release 17,20P into the water and that a rapid (within 15 min) elevation in blood gonadotropin of males mediates the milt response to 17,20P exposure. We conclude that this pheromone system synchronizes milt production with ovulation.

The goldfish is well suited for studies of pheromone release by females and pheromonal effects on males because ovulation is synchronized with the light-dark cycle and can be manipulated under laboratory conditions. Goldfish which have completed

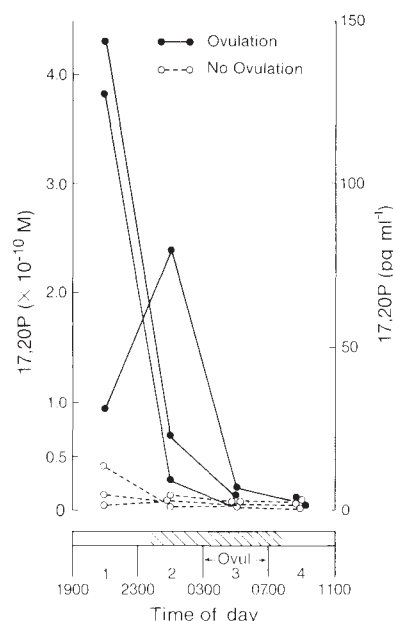


Fig. 1 Concentrations of 17,20P in female holding water during the peri-ovulatory period.

Methods. At 21:00 h on day 1, six mature female goldfish (Ozark Fisheries, Stoutland, Missouri) were transferred from stock tanks (14 °C; 16L:8D photoperiod, lights on at 08:00 h) to aquaria (20 °C) containing aquatic vegetation. At 19:00 h on day 2, when the preovulatory surge of GtH (ref. 9) should have commenced, females were removed and placed individually in glass jars containing 1 litre of aerated water at 20 °C. At 23:00 h, a 10 ml water sample was taken from each jar for 17,20P determination and each female was transferred to a clean jar. This procedure was repeated four times, the last water sample being taken at 11:00 h on day 3. Three of the six females ovulated, all during the third sample period (03:00-07:00 h). Water samples were extracted twice with 2 vol ether and the amount of 17,20P in the extract determined by radioimmunoassay^{15,16}.

vitellogenesis at 12-14 °C ovulate spontaneously within a day of being warmed to 20 °C and exposed to aquatic vegetation, the substrate for oviposition. On a 16 hours light:8 hours dark photoperiod, the preovulatory gonadotropin (GtH) surge begins after the midpoint of the photophase and induces ovulation (follicular rupture) about 10-12 hours later, during the latter half of the scotophase⁹; oocyte final maturation is completed halfway between the initiation of the GtH increase and ovulation¹⁰. Spawning begins shortly after ovulation.

To determine whether or not females release 17,20P, six mature females were maintained individually in 1 litre of water for four successive four-hour intervals encompassing the peri-ovulatory period. Three of the six females ovulated during the third four-hour interval and released high levels of 17,20P into the water during the first and second four-hour intervals (Fig. 1). All water samples from the three females which did not ovulate had relatively low concentrations of 17,20P.

We next used hypophysectomized males to determine whether exposure to 17,20P increases milt volume by stimulating the pituitary-gonadal axis. When hypophysectomized males were stripped of milt¹¹ two days after surgery and then exposed to a water concentration of 17,20P (5×10^{-10} M) approximating the peak concentrations released by ovulating females (Fig. 1), milt volumes which could be stripped the following morning were significantly ($P < 0.001$) reduced. In contrast, milt volumes of males which had undergone a sham operation and of intact males were significantly ($P < 0.001$) increased following this exposure to 17,20P (Fig. 2).

We determined whether GtH, known to increase milt production when injected into a variety of teleosts including goldfish¹²,

Fig. 2 Effect of hypophysectomy on the 17,20P-induced increase in milt volume in male goldfish. Open bars, initial milt volume; shaded bars, milt volume after exposure to 17,20P. Males were hypophysectomized (Hypox.), had undergone a sham operation (Sham) or were intact (Intact); ***, significantly different from pre-exposure values ($P < 0.001$); $n = 20$ in all groups.

Methods. As in other experiments (Figs 3 and 4), mature male goldfish were maintained before use in flow-through stock tanks at 20 °C on a 16L:8D photoperiod (lights on at 08:00 h). On day 1, males were anaesthetized (0.1% 2-phenoxyethanol, Syndel), then either hypophysectomized by the opercular approach¹⁷, sham operated, or handled without surgery, and placed by groups in 65 l flow-through aquaria (four fish per aquarium) at 20 °C. At 17:00 h on day 3, fish were anaesthetized and stripped of milt to determine initial milt volumes. The fish were placed belly-up in a slotted foam pad and gentle finger pressure was applied to the abdomen to express milt into weighed (± 0.1 mg) haematocrit tubes with weighed caps which were then weighed to determine the weight of milt. Milt density was assumed to be 1.0 and milt values are expressed as volumes (μ l) rather than as weights. At 21:00 h, water flow to all aquaria was shut off and 10 μ g 17,20P (Sigma) in 0.1 ml of ethanol was added to each aquarium; this dose should have produced a water concentration of ~ 150 pg ml⁻¹ (5×10^{-10} M), although actual water concentrations were not measured. At 09:00 h on day 4, all fish were again anaesthetized and stripped to determine milt volumes following exposure to 17,20P. Within each group, milt volumes before and after 17,20P exposure were compared by the Wilcoxon matched-pairs signed ranks test.

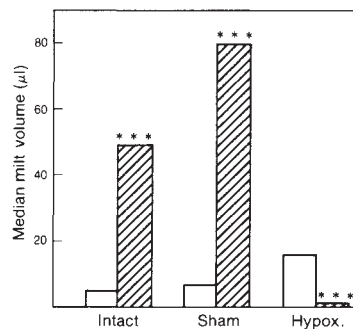
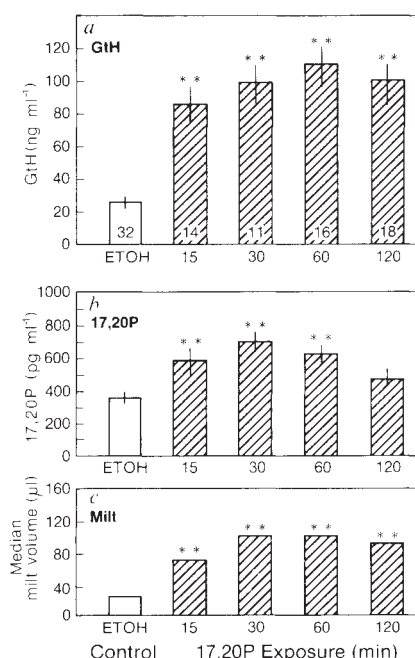


Fig. 3 Effect of water-borne 17,20P on serum GtH (a), serum 17,20P (b), and (c) milt volume in male goldfish. Values in a and b are means; lines, $\pm$ s.e.m. range; **, significantly different from controls ($P < 0.01$). Numbers within bars in a show sample size.

Methods. Mature males were placed in 65 l flow-through aquaria (four fish per aquarium) at 20 °C. During scotophase, 10 μ g of 17,20P in 0.1 ml of ethanol was added to each experimental aquarium and blood samples were taken from different groups of males under anaesthesia 15, 30, 60, and 120 min later; control males were bled following 60 min exposure to ethanol vehicle. All blood samples were taken during the scotophase (02:00–06:00 h) and all treatment groups were balanced over time. Following blood sampling, all fish were transferred to clean flow-through aquaria and milt sampled 7–8 h after exposure to 17,20P had begun. Serum GtH (ref. 18) and 17,20P (refs 15,16) concentrations were determined by radioimmunoassay. Data on GtH and 17,20P were analysed by ANOVA and Newman-Keuls procedure. Milt volume data were analysed by Kruskal-Wallis ANOVA and Mann-Whitney *U* tests.



increases in the blood of males exposed to water-borne 17,20P. We also measured serum 17,20P levels as 17,20P of testicular origin has been shown to mediate the action of GtH on milt production¹³. Experimental males exposed during the scotophase to 17,20P (5×10^{-10} M) for 15, 30, 60, or 120 min had significantly ($P < 0.01$) higher blood GtH concentrations than did control males exposed to the ethanol vehicle for 60 min (Fig. 3a). As well, milt volumes stripped 7–8 h after 17,20P exposure were significantly ($P < 0.01$) greater in the experimental groups than in the control group (Fig. 3c). Consistent with the hypothesis that endogenous 17,20P mediates GtH-induced milt increase in goldfish, serum 17,20P levels in three of the four groups exposed to water-borne 17,20P were significantly ($P < 0.01$) higher than in control males exposed to ethanol (Fig. 3b). It is unlikely that passive uptake of 17,20P from the aquarium water accounts for these differences between experimental and control groups because serum 17,20P levels in all groups were higher than the calculated water concentration of 17,20P.

A pheromone-mediated increase in milt volume could affect

reproductive success if the increase occurs prior to ovulation and spawning. We therefore determined the latency of the 17,20P-induced milt response. Males exposed to 5×10^{-10} M 17,20P or ethanol at the onset of the scotophase were anaesthetized either three or six hours later for blood sampling and determination of milt volume. At both sample times, serum GtH levels in males exposed to water-borne 17,20P were significantly higher ($P < 0.01$ and $P < 0.05$, respectively) than those of control males (Fig. 4a). However, only at the 6 h sample time were the milt volumes of males exposed to 17,20P greater ($P < 0.05$) than those of controls (Fig. 4b). As peak preovulatory release of 17,20P from females occurs more than 4 h prior to ovulation (Fig. 1), the milt response to water-borne 17,20P appears to be rapid enough to allow an increase in milt volume prior to spawning.

The results presented in this and our earlier studies^{2,3} clearly show that 17,20P, which promotes oocyte final maturation in goldfish⁴ and is released from females prior to ovulation, also increases blood GtH and milt volume in males. Indirect evidence that these physiological responses of males are normally stimu-

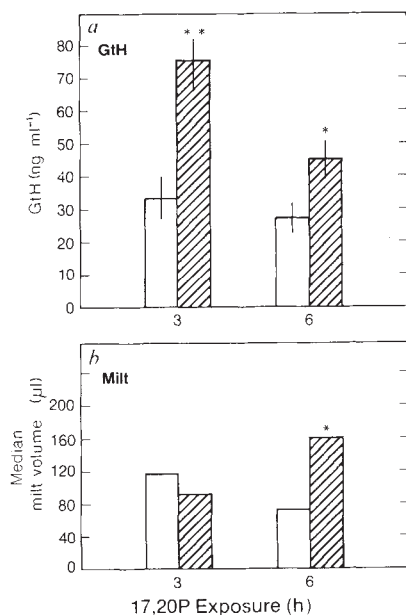


Fig. 4 Effect of 3 and 6 hours exposure to water-borne 17,20P on serum GtH (a) and milt volume (b) in male goldfish. Values in a are means; lines, $\pm$ s.e.m. range; **, $P < 0.01$; *, $P < 0.05$ (significantly different from controls); $n = 15$ in all groups.

Methods. Mature males were placed in 65 l aquaria (three fish per aquarium) at 20 °C. At the onset of scotophase, either 0.1 ml ethanol (open bars) or 10 μ g of 17,20P in 0.1 ml of ethanol was added to each aquarium. Three and 6 h later, different groups of males were anaesthetized for blood sampling and determination of milt volume. Data on GtH were analysed by *t*-tests and milt volume data were analysed by Mann-Whitney *U* tests.

lated by females is provided by the recent observation¹⁴ that the presence of an ovulating female goldfish induces a GtH surge in males coincident with that of the female. Both the temporal pattern of 17,20P release from preovulatory females and the latency of the milt response to water-borne 17,20P support our proposal that 17,20P is a pheromone which synchronizes milt production with ovulation. This appears to be the only vertebrate system in which the identity, cellular source and release of the pheromone, as well as the sensory pathway, endocrine mediators and significance of the response in the pheromone recipient, all are clearly established.

This research was supported by the Alberta Heritage Foundation for Medical Research and the Natural Sciences and Engineering Research Council of Canada.

Received 11 September; accepted 4 November 1986.

1. Stacey, N. E., Kyle, A. L. & Liley, N. R. in *Chemical Communication in Vertebrates* Vol. IV (eds Duvall, D., Muller-Schwarze, D. & Silverstein, R. M.) 117-133 (Plenum, New York, 1986).
2. Stacey, N. E. & Sorensen, P. W. *Can. J. Zool.* (in the press).
3. Sorensen, P. W., Hara, T. J. & Stacey, N. E. *J. comp. Physiol.* (in the press).
4. Nagahama, Y. *et al. Gen. comp. Endocr.* **51**, 15-23 (1983).
5. Goetz, F. W. in *Fish Physiology*, Vol. IXB (eds Hoar, W. S., Randal, D. J. & Donaldson, E. M.) 117-170 (Academic, New York, 1983).
6. Colombo, L., Belvedere, P. C., Marconato, A. & Bentivegna, F. in *Proc. Int. Symp. Reprod. Physiol. Fish.* (eds Richter, C. J. J. & Goos, H. J. Th.) 84-94 (Pudoc, Wageningen, 1982).
7. Van den Hurk, R. & Lambert, J. G. D. *Can. J. Zool.* **61**, 2381-2387 (1982).
8. Stacey, N. E. & Kyle, A. L. *Physiol. Behav.* **30**, 621-628 (1983).
9. Stacey, N. E., Cook, A. F. & Peter, R. E. *Gen. comp. Endocr.* **37**, 246-249 (1979).
10. Stacey, N. E. *et al. Can. J. Zool.* **61**, 2646-2652 (1984).
11. Kyle, A. L., Stacey, N. E., Peter, R. E. & Billard, R. *Gen. comp. Endocr.* **57**, 10-22 (1985).
12. Billard, R., Postier, A., Weil, C. & Breton, B. *Can. J. Fish. aquat. Sci.* **39**, 65-79 (1982).
13. Ueda, H., Kambe-gawa, A. & Nagahama, Y. *Gen. comp. Endocr.* **59**, 24-30 (1985).
14. Kobayashi, M., Aida, K. & Hanyu, I. *Gen. comp. Endocr.* **62**, 70-79 (1986).
15. Scott, A. P., Sheldrick, E. L. & Flint, A. P. F. *Gen. comp. Endocr.* **46**, 444-451 (1982).
16. Van Der Kraak, G., Dye, H. M. & Donaldson, E. M. *Gen. comp. Endocr.* **55**, 36-45 (1984).
17. Yamazaki, F. *Mem. Fac. Fish. Hokkaido Univ.* **13**, 1-65 (1965).
18. Peter, R. E., Nahorniak, C. S., Chang, J. P. & Crim, L. W. *Gen. comp. Endocr.* **55**, 337-346 (1984).

Astrocytes induce blood-brain barrier properties in endothelial cells

Robert C. Janzer* & Martin C. Raff

Medical Research Council Developmental Neurobiology Programme, Zoology Department, University College London, London WC1E 6BT, UK

The highly impermeable tight junctions between endothelial cells forming the capillaries and venules in the central nervous system (CNS) of higher vertebrates are thought to be responsible for the blood-brain barrier^{1,2} that impedes the passive diffusion of solutes from the blood into the extracellular space of the CNS³⁻⁶. The ability of CNS endothelial cells to form a blood-brain barrier is not intrinsic to these cells but instead is induced by the CNS environment: Stewart and Wiley⁷ demonstrated that when avascular tissue from 3-day-old quail brain is transplanted into the coelomic cavity of chick embryos, the chick endothelial cells that vascularize the quail brain grafts form a competent blood-brain barrier; on the other hand, when avascular embryonic quail coelomic grafts are transplanted into embryonic chick brain, the chick endothelial cells that invade the mesenchymal tissue grafts form leaky capillaries and venules. It is, however, not known which cells in the CNS are responsible for inducing endothelial cells to form the tight junctions characteristic of the blood-brain barrier. Astrocytes are the most likely candidates since their processes form endfeet that collectively surround CNS microvessels^{8,9}. In this report we provide direct evidence that astrocytes are capable of inducing blood-brain barrier properties in non-neural endothelial cells *in vivo*.

Astrocytes were purified from neonatal Sprague-Dawley (S/D) rat cerebral cortex as previously described¹⁰. More than 95% of these cells were labelled by an antiserum against glial fibrillary acidic protein (GFAP), a useful marker for astrocytes in the CNS¹¹. All of the GFAP-positive cells had a morphological and antigenic phenotype characteristic of type I astrocytes¹². When a suspension of these cells was injected into the anterior eye chamber of adult syngeneic rats, as described by Olson *et al.*¹³, the astrocytes formed aggregates on the surface of the iris and these became vascularized within 48 hours, as observed with a dissecting microscope (not shown). The dye Evans blue, which binds strongly to endogenous albumin¹⁴, was injected intravenously 2 weeks after the injection of the astrocytes in order to assess the 'tightness' of the newly formed vessels in the astrocyte aggregates: although most of the tissues in the rat were stained blue, the astrocyte aggregates remained white (Fig. 1A), suggesting that the capillaries and venules in the aggregates were tight. When meningeal cells, prepared from neonatal rats as previously described¹⁵, were injected instead of astrocytes they also formed vascularized aggregates on the surface of the iris, but these stained blue following an injection of Evans blue (Fig. 1B), indicating that capillaries and venules in the aggregates were leaky.

These macroscopic findings were confirmed microscopically in frozen sections. Following Evans blue injection, a large amount of the dye, which fluoresces red under UV light¹⁶, was found by fluorescence microscopy in the extracellular space in sections of meningeal cell aggregates but not in sections of astrocyte aggregates, even though both types of aggregates contained small blood vessels (not shown). Similarly, when endogenous albumin was stained immunohistochemically, intense labelling was seen in the extracellular space in sections of meningeal cell aggregates (Fig. 3D), but not in sections of astrocyte aggregates (Fig. 3B). Thus endothelial cells formed capillaries and venules that were functionally tight in astrocyte aggregates and functionally leaky in meningeal cell aggregates. As expected, in the former aggregates, GFAP-positive astrocyte

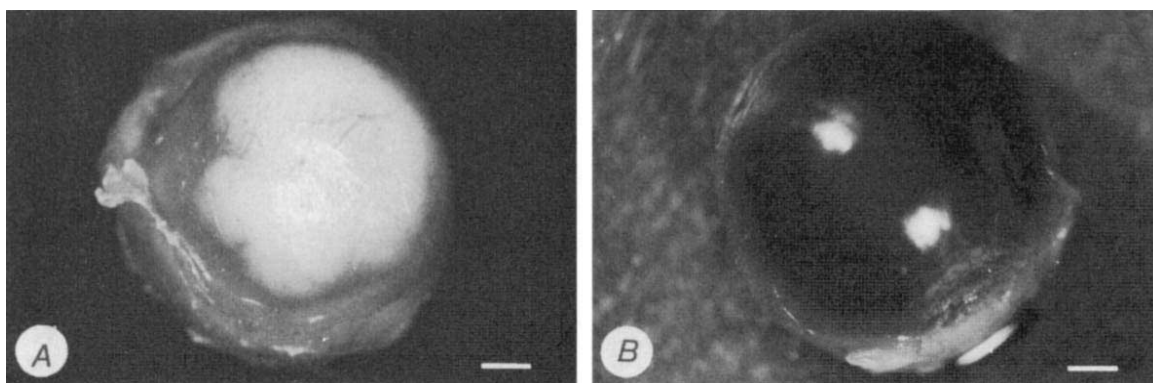


Fig. 1 Vascularized aggregates of astrocytes (*A*) and meningeal cells (*B*) in the anterior eye chamber after injection of Evans blue. The astrocyte cell aggregate in *A* is unstained; the meningeal cell aggregate in *B* is intensely stained. The two white patches in *B* are optical artefacts. Scale bars, 1 mm.

Methods. Rat astrocytes were prepared as previously described¹⁰, using a modification of the procedure of McCarthy and de Vellis²¹ as described by Noble *et al.*²². In brief, cerebral cortex cultures from 1–2 day old S/D rats were grown in Falcon flasks for 7–9 days in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (DMEM-FCS). The flasks were shaken overnight on a rotatory platform to remove the top cells, mainly neurons and cells of the oligodendrocyte-type-2-astrocyte (O-2A) lineage²³. The cultures were then treated with cytosine arabinoside (2×10^{-5} M) for 2 days to kill rapidly dividing cells. The cells were then removed from the culture flask with 0.125% (w/v) trypsin and 0.02% (w/v) EDTA (trypsin-EDTA) and recultured in flasks in DMEM-FCS for 7–14 days, after which they were removed with trypsin-EDTA, washed in DMEM-FCS and suspended in DMEM for injection. More than 95% of the cells were GFAP-positive²⁴. In some cases, before reculturing, the cells were treated in suspension with monoclonal A2B5²⁵ and anti-galactocerebroside (GC)²⁶ antibodies in the presence of rabbit complement as previously described¹² to kill any remaining cells of the O-2A lineage; less than 1% of the cells in such preparations were A2B5-positive or GC-positive. The results obtained were similar with antibody-treated and untreated astrocytes. Meningeal cells were prepared from neonatal S/D rats as previously described¹⁵ and, after five passages, were prepared for injection as described for astrocytes. Recipient rats were anaesthetised with ether and their pupils dilated with 1% atropine-sulphate applied locally. A no. 27 gauge needle was inserted into the posterior eye chamber and left in place during the injection procedure to allow the intraocular pressure to equilibrate. The cells ($0.5\text{--}1 \times 10^6$ suspended in 5 μ l of DMEM) were injected slowly into the anterior eye chamber through a glass pipette connected via plastic tubing to a motor-driven microinjection device. Two weeks later, the rats were reanaesthetised and 1 ml of 2% (w/v) Evans blue (Sigma) in Sorensen buffer pH 7.2 was injected intravenously. After 2–10 min, the rats were perfused through the ascending aorta with 4% paraformaldehyde in the same buffer containing 7% (w/v) sucrose. The eyes were removed and immersion-fixed in the same fixative. Photographs were taken with a Nikon dissecting photomicroscope using Ilford PF4 film.

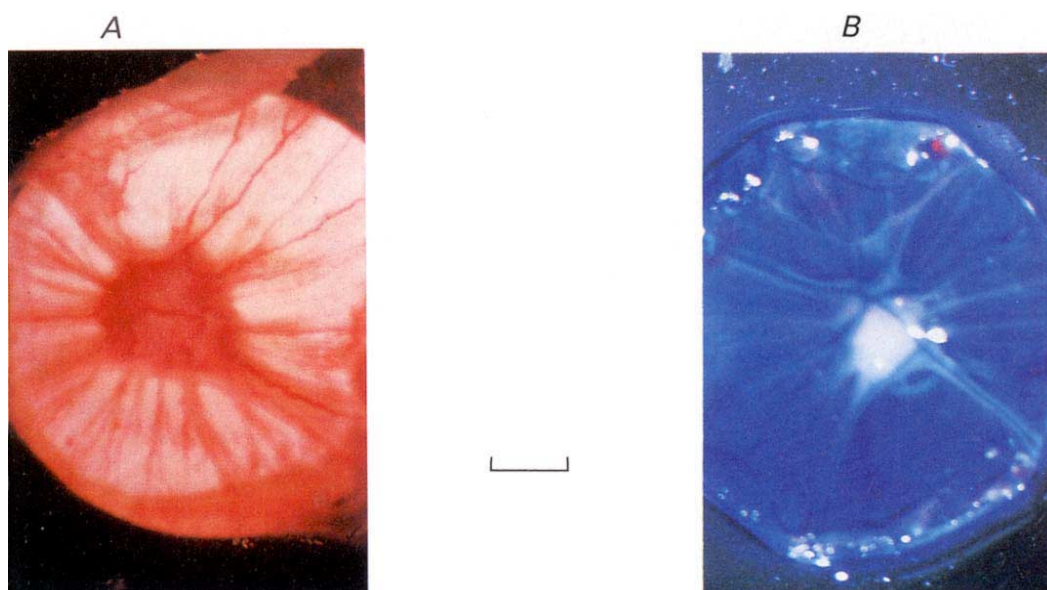


Fig. 2 Vascularized astrocyte aggregates growing on Millipore filters inverted on chick chorioallantoic membranes, *A*, before and *B*, after Evans blue perfusion. Scale bar, 2 mm.

Methods. Astrocytes (5×10^5 cells), prepared as described in Fig. 1, were cultured in a small spot on a Millipore filter (pore size $0.45 \mu\text{m}$) for 24 h in DMEM-FCS. The filter was then put cell-side down onto the chorioallantoic membrane of a 5-day-old chick embryo²⁹. The egg was sealed with parafilm and reincubated for 11 days. In *A*, the filter with the adherent chorioallantoic membrane and astrocyte aggregate was cut out and photographed. In *B*, the embryo was perfused with Evans blue, as described in Fig. 1, before the filter with adherent tissue was cut out and photographed.

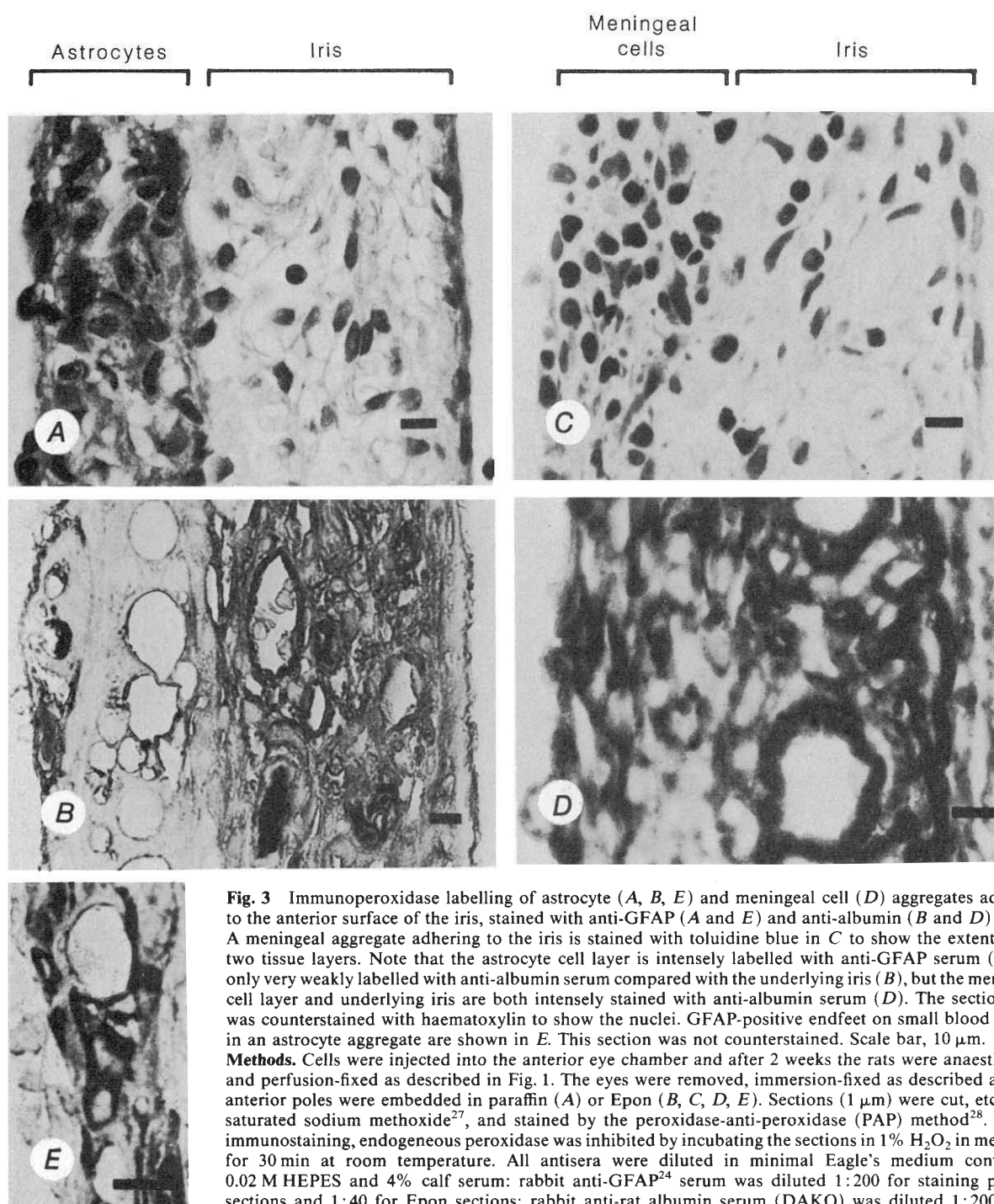


Fig. 3 Immunoperoxidase labelling of astrocyte (A, B, E) and meningeal cell (D) aggregates adhering to the anterior surface of the iris, stained with anti-GFAP (A and E) and anti-albumin (B and D) serum. A meningeal aggregate adhering to the iris is stained with toluidine blue in C to show the extent of the two tissue layers. Note that the astrocyte cell layer is intensely labelled with anti-GFAP serum (A) but only very weakly labelled with anti-albumin serum compared with the underlying iris (B), but the meningeal cell layer and underlying iris are both intensely stained with anti-albumin serum (D). The section in A was counterstained with haematoxylin to show the nuclei. GFAP-positive endfeet on small blood vessels in an astrocyte aggregate are shown in E. This section was not counterstained. Scale bar, 10 μ m.

Methods. Cells were injected into the anterior eye chamber and after 2 weeks the rats were anaesthetised and perfusion-fixed as described in Fig. 1. The eyes were removed, immersion-fixed as described and the anterior poles were embedded in paraffin (A) or Epon (B, C, D, E). Sections (1 μ m) were cut, etched in saturated sodium methoxide²⁷, and stained by the peroxidase-anti-peroxidase (PAP) method²⁸. Before immunostaining, endogenous peroxidase was inhibited by incubating the sections in 1% H_2O_2 in methanol for 30 min at room temperature. All antisera were diluted in minimal Eagle's medium containing 0.02 M HEPES and 4% calf serum: rabbit anti-GFAP²⁴ serum was diluted 1:200 for staining paraffin sections and 1:40 for Epon sections; rabbit anti-rat albumin serum (DAKO) was diluted 1:200. Both rabbit antisera were visualized with sheep anti-rabbit immunoglobulin (DAKO, 1:40) followed by rabbit

anti-peroxidase complexed with peroxidase (PAP) (DAKO, 1:100) and then 3,3'-diaminobenzidine (DAB) containing 0.005% H_2O_2 ; the staining was intensified with 2% OsO_4 . All incubations were at room temperature for 60 min, followed by three 20-min washes in 0.1 M phosphate buffer, containing 4% calf serum and 0.1% Triton-X-100. The section in A was counterstained with haematoxylin and that in C was stained with Toluidine blue. Sections were mounted in glycerol-gelatin and photographed with a Zeiss Universal microscope using Ilford PF4 film. In control experiments, no peroxidase labelling was seen when the rabbit antisera were replaced by normal rabbit serum at the same dilution.

endfeet were seen around many small blood vessels (Fig. 3E).

The most likely explanation for these findings is that astrocytes induced invading, non-neural endothelial cells to form tight capillaries and venules, but an alternative explanation is that the tight vessels in the astrocyte aggregates were formed by neural endothelial cells that either contaminated the purified astrocyte inoculum or were specifically attracted from the neural retina by the astrocytes. To exclude the latter possibility, rat

astrocytes were prepared as above and then cultured at high density in a small area on a Millipore filter for 24 hours. The filters were then placed (cells down) onto the exposed chorioallantoic membrane of 5-day-old chick embryos. After 11 days the chick embryos were injected intravenously with Evans blue: the chorioallantoic membrane turned blue, but the tissue lump formed by the astrocytes remained white (Fig. 2). On the other hand, when filters containing embryonic rat limb fibroblasts

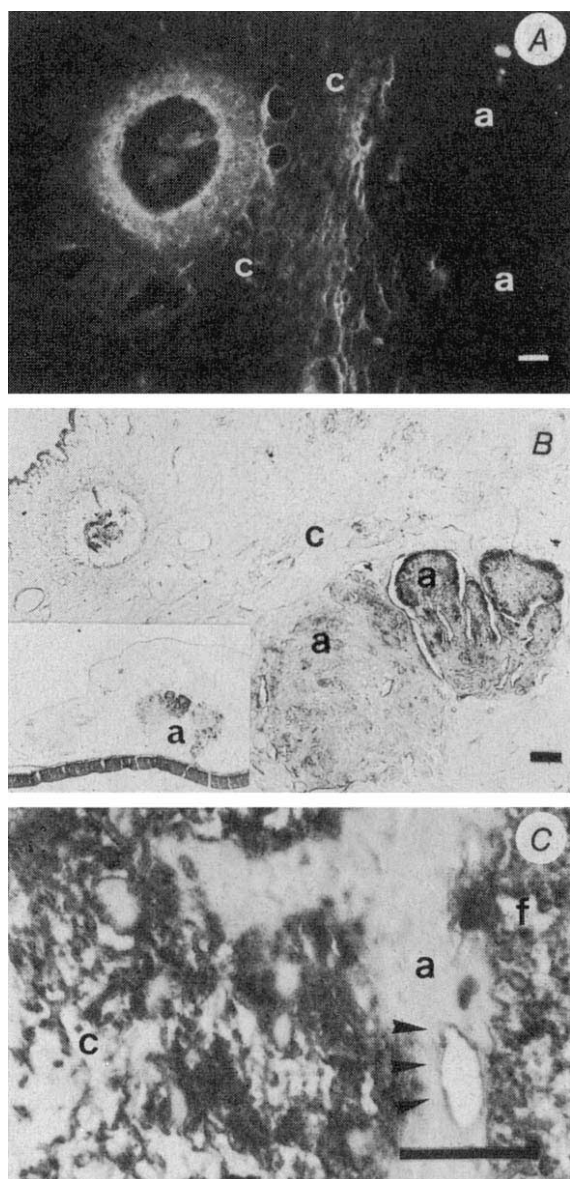


Fig. 4 Fluorescence (A) and immunoperoxidase (B and C) micrographs of astrocyte aggregates growing on Millipore filters inverted on chick chorioallantoic membranes, after Evans blue perfusion (A) or staining with rabbit anti-GFAP serum (B) or rat anti-chick-cell serum (C); A and B are sections of the same block which is shown as an inset in B: the astrocyte aggregate (marked with 'a'), which is GFAP-positive (B), is largely unstained by Evans blue, as shown by the lack of fluorescence (A), but the underlying chorioallantoic membrane (marked with 'c'), which is GFAP-negative (B), is stained by Evans blue, as indicated by the bright fluorescence (A). In C, all cells in the chorioallantoic membrane are stained with anti-chick-cell serum, while only endothelial cells (marked with arrowheads) are stained within the astrocyte aggregate; the underlying Millipore filter (marked with 'f') is non-specifically stained. Scale bar, 20 μ m.

Methods. Astrocytes were grown on Millipore filters inverted on a chick chorioallantoic membrane as described in Fig. 2. After 11 days the filter with the adherent chorioallantoic membrane and astrocyte aggregate was cut out and immersion-fixed at room temperature for 30 min as described in Fig. 1. In A, the embryo was perfused with Evans blue before the filter was cut out. The area containing the astrocyte aggregate was cut out and either frozen in isopentane precooled in liquid nitrogen followed by immersion in liquid nitrogen (A and B) or embedded in Epon (C). In (A) and (B), cryostat sections (10 μ m) were cut and thawed onto poly-L-lysine coated glass slides. Immunoperoxidase staining was performed using rabbit anti-GFAP serum (diluted 1:100) as described in Fig. 2 (on cryostat sections the endogenous peroxidase blocking step was omitted). For C, an antiserum against embryonic chick cells was produced by injecting mechanically-dissociated cells from the liver, spleen, muscles and kidneys of 18-day chick embryos into adult S/D rats. The cells were emulsified with an equal volume of complete Freund's adjuvant and injected both subcutaneously and intraperitoneally. The rats were immunised again in the same way 2 weeks after the first immunisation and were bled 7 days later. When diluted 1:50, this antiserum (using the method described below) stained frozen sections of chick but not rat brain. The antiserum was applied, diluted 1:10 in MEM-HEPES containing 50% (v/v) normal chick serum, to etched, 1- μ m Epon sections (prepared as described in Fig. 3). Antibody binding was visualised by the avidin-biotin-complex (ABC) method¹⁷ using a biotinylated rabbit anti-rat-IgG antibody (Vector Laboratories, diluted 1:50) followed by horseradish-peroxidase conjugated ABC (Vector Laboratories). DAB-staining was performed as described in Fig. 2. In all immunostaining experiments, no staining was observed when normal sera were used at the same dilutions as the antisera.

were placed on chorioallantoic membranes, these cell aggregates turned blue after Evans blue injection (not shown). When frozen sections of the astrocyte and fibroblast aggregates were examined by fluorescence microscopy, large amounts of extracellular dye were seen in the chorioallantoic membrane and in the fibroblast aggregates but not in the astrocyte aggregates (Fig. 4A). Similar results were obtained when endogenous albumin in such frozen sections was labelled by indirect immuno-fluorescence with a rabbit antiserum against chick albumin (not shown). To exclude the possibility that the vessels within the astrocyte aggregates growing on chorioallantoic membranes were formed by neural endothelial cells contaminating the cultured astrocyte population, Epon-embedded, etched, 1- μ m sections of such aggregates were labelled by the ABC immunoperoxidase method¹⁷ using an antiserum raised against dissociated chick embryo cells. Endothelial cells within the aggregate were intensely stained by the antiserum whereas the astrocytes were not (Fig. 4C), indicating that the astrocyte aggregates were vascularized by host-derived chick endothelial cells, rather than by contaminating rat endothelial cells.

Our results provide direct evidence that astrocytes (but not meningeal cells or fibroblasts) can induce endothelial cells of

non-nervous system origin to form non-leaky vessels, strongly suggesting that these glial cells are responsible for inducing capillary and venule endothelial cells in the CNS to form the blood-brain barrier. The molecules that mediate this cell-cell interaction must have been highly conserved as they operate between chick and rat cells. In order to characterize these molecules it will probably be necessary to develop *in vitro* models of this interaction. In this regard it is promising that astrocytes can apparently enhance the frequency, length and complexity of tight junctions formed between brain-derived endothelial cells in culture¹⁸ and that cultured glial tumor cells can induce cells of a brain endothelial cell line to express an enzyme¹⁹ and polarized transport²⁰ characteristic of normal brain endothelial cells.

We thank E. Clarke, B. Walker and L. Kläusli for excellent technical help, Dr M. Ratcliffe for help with producing the rat anti-chick-cell serum, and our colleagues in the MRC Developmental Neurobiology Programme for helpful comments and encouragement. Preliminary experiments by Drs B. Williams and S. David helped guide the work described in this report. R.C.J. was supported by a grant from the Educational Council of the Canton of Zurich (Switzerland).

Received 16 October; accepted 13 November 1986.

1. Reese, T. S. & Karnovsky, M. J. *J. Cell Biol.* **34**, 207-217 (1967).
2. Nagy, Z., Peters, H. & Huettner, I. *Lab. Invest.* **50**, 313-322 (1984).
3. Crone, C. *Acta physiol. Scand.* **64**, 407-417 (1965).
4. Bradbury, M. W. B. *Circulation Res.* **57**, 213-222 (1985).
5. Abbott, N. J. *Trends Neurosci.* **9**, 3-7 (1986).
6. Betz, A. L. & Goldstein, G. W. A. *Rev. Physiol.* **48**, 241-250 (1986).
7. Stewart, P. A. & Wiley, M. J. *Dev. Biol.* **84**, 183-192 (1981).
8. Ramon y Cajal, S. in *Histologie du système nerveux de l'homme et des vertébrés* Vol. 1, 230-252, 2nd edn (CSDIC, Instituto Ramon y Cajal, Madrid, 1972).
9. Peters, A., Palay, S. L. & DeWebster, H. in *The Fine structure of the nervous system: The neurons and supporting cells* 231-263 (Saunders, Philadelphia, 1976).
10. Raff, M. C., Abney, E. R. & Fok-Seang, J. *Cell* **42**, 61-69 (1985).
11. Bignami, A., Eng, L. F., Dahl, D. & Uyeda, C. T. *Brain Res.* **43**, 429-435 (1972).
12. Raff, M. C., Abney, E. R., Cohen, J., Lindsay, K. & Noble, M. J. *Neurosci.* **3**, 1289-1300 (1983).
13. Olson, L., Seiger, A. & Stroemberg, I. in *Advances in Cellular Neurobiology* Vol. 4 (eds Federoff, S. & Hertz, L.) 407-422 (Academic, New York, 1983).
14. Freedman, F. B. & Johnson, J. A. *Am. J. Physiol.* **216**, 675-681 (1969).
15. Raff, M. C. *et al. Brain Res.* **174**, 283-308 (1979).
16. Saria, A. & Lundberg, J. M. *J. Neurosci. Meth.* **8**, 41-49 (1983).
17. Hsu, S. M., Raine, L. & Fanger, H. *J. Histochem. Cytochem.* **29**, 577-584 (1981).
18. Tao-Cheng, J.-H., Nagy, Z. & Brightman, W. *Anat. Rec.* **214**, 131A (1986).
19. DeBault, L. E. & Cancilla, P. A. *Science* **207**, 653-655 (1980).
20. Beck, D. W., Vinters, H. V., Hart, M. N. & Cancilla, P. A. *J. Neuropath. exp. Neurol.* **43**, 219-224 (1984).
21. McCarthy, K. D. & de Vellis, J. *J. Cell Biol.* **85**, 890-902 (1980).
22. Noble, M., Fok-Sang, J. & Cohen, J. *J. Neurosci.* **4**, 1892-1903 (1984).
23. Raff, M. C., Abney, E. R. & Miller, R. H. *Dev. Biol.* **106**, 53-60 (1984).
24. Pruss, R. M. *Nature* **280**, 688-690 (1979).
25. Eisenbarth, G. S., Walsh, F. S. & Nirenberg, M. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4913-4917 (1979).
26. Ranscht, B., Clapshaw, P. A., Price, J., Noble, M. & Siefert, W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2707-2713 (1982).
27. Pino, R. M. & Thouron, C. L. *J. Histochem. Cytochem.* **31**, 411-416 (1983).
28. Sternberger, L. A. *Immunocytochemistry*, 3rd edn, (Wiley, New York, 1986).
29. New, D. A. T. *The Culture of Vertebrate Embryos* 71-73 (Academic, London, 1966).

Capillary endothelial cells express basic fibroblast growth factor, a mitogen that promotes their own growth

Lothar Schweigerer*, Gera Neufeld*, Jeff Friedman†, Judith A. Abraham†, John C. Fiddes† & Denis Gospodarowicz*

*Cancer Research Institute (M-1282), University of California Medical Center, San Francisco, California 94143, USA

†California Biotechnology Inc., 2450 Bayshore Parkway, Mountain View, California 94043, USA

Angiogenesis, the formation of new capillaries, which is observed in embryonic and injured tissue and is particularly prominent in the vicinity of solid tumours¹, involves the migration and proliferation of capillary endothelial cells. It is probably triggered by agents, such as basic fibroblast growth factor (bFGF), thought to be released from tissues adjacent to proliferating capillaries¹. As well as being a potent inducer of cell division in capillary endothelial cells *in vitro*, bFGF can act as an angiogenic agent *in vivo*². It is present in a wide variety of richly vascularized tissues including brain, pituitary, retina, adrenal gland, kidney, corpus luteum, placenta and various tumours¹⁻³. So far, however, the normal bFGF-producing cell species in these tissues have not been identified^{2,4}. We report here that capillary endothelial cells express the bFGF gene, that they produce and release bFGF and that bFGF derived from them can stimulate the proliferation of capillary endothelial cells. We conclude that bFGF can act as a self-stimulating growth factor for capillary endothelial cells, and that it is possible that the formation of new capillaries is induced by capillary endothelial cells themselves.

Initially we found that crude extracts prepared from cultured bovine adrenal-cortex-derived or brain-derived capillary endothelial cells (ACE or BCE cells, respectively)⁵, but not from the human leukaemia cell lines HUT-78 or JURKAT⁶, stimulate capillary endothelial cell proliferation (data not shown). This suggested that the capillary endothelial cells, but not control leukaemia cells, might contain bFGF.

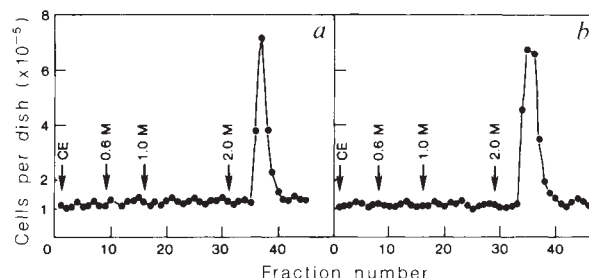


Fig. 1 Heparin-Sepharose (HS) affinity chromatography of crude extracts of (a) ACE or (b) BCE cells. ACE or BCE cells were maintained in stock plates containing Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% calf serum (CS) and antibiotics, and received bFGF (1 ng ml⁻¹) every other day⁵. The cells were passaged continuously as described⁵ and used between passages 5 and 16. Confluent cultures were washed with phosphate-buffered saline (PBS), trypsinized, and the cells were centrifuged (300 g, 5 min, 21 °C). The pellet was washed again with PBS, and 3 × 10⁷ cells were suspended in 2 ml of 10 mM Sodium phosphate buffer (pH 7.4) containing 0.15 M NaCl. Cells were lysed by rapid freezing and thawing followed by passage through a 26 gauge needle. After the NaCl concentration was adjusted to 0.2 M, nuclei and cell debris were pelleted by centrifugation (50,000 g, 10 min, 4 °C) and the clear supernatant (referred to as crude cell extract, CE) was applied to an HS column (bed volume, 0.8 ml) that had been equilibrated with 10 mM Tris-HCl buffer, pH 7.0/0.6 M NaCl. The column was washed with the same buffer and eluted sequentially with the same buffer that contained 1.0 M or 2.0 M NaCl (refs 8-10), as indicated. The flow rate was 15.6 ml h⁻¹. Fractions (0.52 ml) were collected and assayed for their abilities to stimulate cell growth as follows: fractions were diluted two-fold with DMEM containing 0.5% BSA and 10 µl aliquots of the diluted samples were added every other day to BCE cells that had been seeded at a density of 2 × 10⁴ per dish. After four days, cells were trypsinized and counted in a Coulter particle counter⁵. Values are the means of triplicate determinations which varied by less than 10%.

Subsequently, crude extracts were subjected to heparin-Sepharose (HS) affinity chromatography⁷, for rapid and selective purification of FGF^{3,8}. All the mitogenic activity present in the crude ACE or BCE cell extracts was retained by HS columns; subsequent elution with 0.6 M, 1.0 M and 2.0 M NaCl yielded a single peak of bioactivity which in both cases eluted with 2.0 M NaCl (Fig. 1a and b, respectively). This chromatographic behaviour is similar to that reported for bFGF and unlike that of acidic FGF (aFGF), which elutes with 1.0 M NaCl⁸⁻¹⁰. No bioactivity was eluted when crude extracts of the leukaemia cell lines were subjected to HS affinity chromatography (data not shown).

To investigate further the possibility that the capillary-endothelial-cell-derived bioactivity that eluted with 2.0 M NaCl might be bFGF, we examined the pooled, bioactive fractions (fractions 5-7) of the 2.0 M NaCl eluates (referred to as purified extracts) for their abilities to inhibit binding of ¹²⁵I-labelled bFGF (¹²⁵I-bFGF) to the FGF receptor¹¹. As shown in Fig. 2a, purified extracts of both ACE and BCE cells inhibited specific ¹²⁵I-bFGF binding in a concentration-dependent manner. The abilities of the purified extracts of both cell types to inhibit binding correlated well with their abilities to stimulate the proliferation of ACE cells (Fig. 2b) or human umbilical vein endothelial cells (not shown), in spite of the fact that the assays employed differ in their sensitivities to bFGF^{2,5,11}. Purified extracts of the leukaemia cell lines neither inhibited specific ¹²⁵I-bFGF binding (Fig. 2a), nor stimulated the proliferation of ACE cells (Fig. 2b). Thus, purified ACE or BCE cell extracts contain a bFGF-like mitogen that stimulates the growth of capillary endothelial cells.

The growth-stimulatory activity of the purified ACE cell extract, like that of native bFGF (ref. 12), was completely

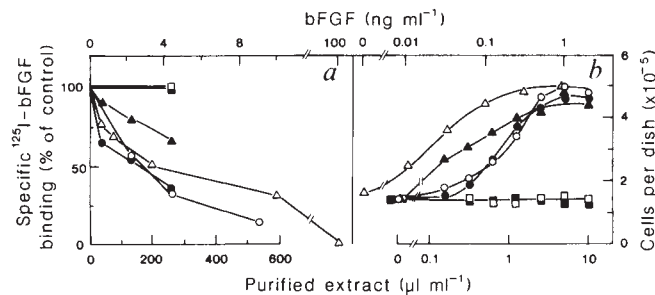


Fig. 2 Effect of HS-purified extracts of capillary endothelial or leukemia cells on (a) specific ¹²⁵I-bFGF binding to baby hamster kidney (BHK-21) cell membranes or on (b) the growth of adrenal-cortex-derived capillary endothelial (ACE) cells. *a*, Purified extracts of BCE cells (▲) or ACE cells grown with (●) or without (○) bFGF, and of the leukaemia cell lines JURKAT (■) or HUT-78 (□) or different concentrations of native pituitary bFGF (ref. 8) (Δ), were examined for their abilities to inhibit specific binding of ¹²⁵I-bFGF (250 pM) to BHK-21 cell membranes (16 μg of protein)¹¹. Specific ¹²⁵I-bFGF binding, which represents bound ¹²⁵I-bFGF that was displaced by an excess (100 ng ml⁻¹) of unlabelled bFGF, was determined as described¹¹. *b*, The indicated amounts of purified capillary endothelial or leukaemia cell extracts or of native bFGF (symbols as in *a*) were added every other day to ACE cells that had been seeded at a density of 2×10^4 per 35 mm dish. After 5 days, cells present in the dishes were counted. The results are representative for a total of 5 separately performed experiments. Values in *a* and *b* represent the means of triplicate determinations which varied by less than 10%.

Methods. Capillary endothelial cells were grown to confluency as described above. Leukaemia cells were grown without bFGF in DMEM/Ham's F-12 (1/1) medium containing 10% CS and antibiotics. Capillary endothelial cells were washed with PBS and trypsinized before extraction. Capillary endothelial or leukaemia cells (3×10^8) were collected, lysed, and the lysate centrifuged. Protein from the resulting supernatants (25 mg) was then subjected to HS affinity chromatography and the fractions eluted were analysed for the presence of bioactive material (details in legend to Fig. 1). If present, bioactive material was always recovered in fractions 5–7 of the 2 M NaCl eluate. The pooled fractions (purified cell extract) were stored at -70°C . Aliquots were dialysed before use in the binding assay.

destroyed by exposure to heat (5 min at 65°C) or acidic pH (pH 2.0, 2 h at 22°C). Furthermore, the growth-stimulatory effect was completely blocked by anti-bFGF antibodies, but not by control antibodies (Fig. 3, solid bars). Thus, the bFGF-like mitogen present in ACE cell extracts is responsible for the growth stimulation of ACE cells.

Although the anti-bFGF antibodies were able to inhibit completely the proliferation induced by the ACE cell extract-derived bFGF, they had only a small, but consistently observed, effect on the basal proliferation of ACE cells that did not receive exogenous growth factor (Fig. 3, open bars). This suggested that most of the bFGF present in the ACE cells might not be released into the culture medium, but might rather act at intracellular sites, a possibility consistent with the demonstration that the bFGF gene does not code a conventional signal peptide sequence¹³. On the other hand, the existence of specific FGF receptors on the surface of a wide variety of cells (refs 11, 14, 15 and unpublished data) indicates that bFGF must reach the cell exterior by some mechanism. Indeed, bFGF binds with high affinity to basement membrane components such as heparan sulphate proteoglycans, which are known to be deposited on the surface of capillary endothelial and other cells¹⁶. Thus, bFGF might be bound to heparan sulphate proteoglycans and as such be transported to the extracellular matrix, where it could be active. However, the association of bFGF with extracellular matrix components might mask its immunoreactive epitopes and

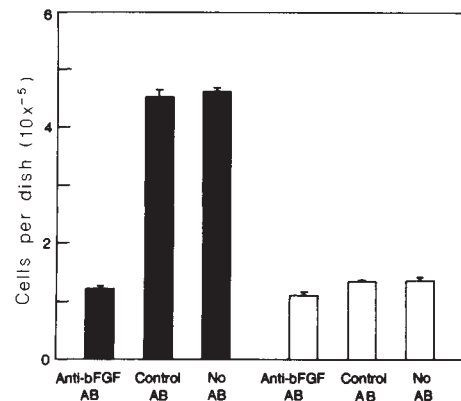


Fig. 3 Effect of specific anti-bFGF antibodies on the proliferation of ACE cells. ACE cells were seeded at a density of 2×10^4 per 35 mm dish and received every other day rabbit anti-bFGF antibodies 'Ede' or rabbit control antibodies (affinity-purified IgG fraction, Flow Laboratories) ($6 \mu\text{g ml}^{-1}$ each) or no antibodies, as indicated, and then either purified ACE cell extract (an equivalent of 1 ng of bFGF ml⁻¹) (solid bars) or no extract (open bars). Values are the means ($\pm$ s.d.) of triplicate determinations. AB, antibodies.

Methods. Rabbit anti-bFGF antiserum obtained by subcutaneous and intracutaneous injection²¹ of pure pituitary-derived bFGF (ref. 8) was purified by protein A-Sepharose affinity chromatography²². The antibodies obtained in the IgG fractions ('Ede') recognize bFGF ($\text{ED}_{50} = 2 \text{ ng ml}^{-1}$), but not aFGF, EGF, insulin, transferrin, cytochrome *c* or thyroglobulin (at concentrations up to $0.4 \mu\text{g ml}^{-1}$) in a specific radioimmunoassay, where ¹²⁵I-bFGF is used as the ligand²³. These antibodies inhibit the bFGF-, but not aFGF-stimulated proliferation of ACE or BCE cells.

thereby prevent full neutralization of its bioactivity by the anti-bFGF antibodies. We found that bFGF is indeed present in the extracellular matrix, so that bFGF might be released from these sites into the culture medium.

To assess whether bFGF can be released from ACE cells, serum-free medium (21) conditioned by long-term cultures of ACE cells was chromatographed over HS and then analysed as outlined in the legend to Fig. 1. The elution profile was similar to that obtained with the ACE cell extract (see Fig. 1), with all of the bioactivity ($\sim 10 \text{ pg}$ of bFGF-equivalents per ml of conditioned medium) being recovered in the 2.0 M NaCl eluate (data not shown). Like native bFGF (ref. 11), the material that was eluted with 2.0 M NaCl inhibited ¹²⁵I-bFGF binding to the FGF receptor and it was able to stimulate the proliferation of ACE cells. Finally, like purified ACE cell extract, the purified conditioned medium lost its ability to stimulate ACE cell growth upon incubation with anti-bFGF antibodies, but not with control antibodies or without antibodies (data not shown). Thus, ACE cells release into the medium a factor indistinguishable from bFGF, that can stimulate their own growth.

The mechanisms by which bFGF is released into the medium are unknown. It is possible, however, that cell lysis or leakage occurring in long-term cultures might be a contributing factor. The existence of similar release mechanisms has been proposed for interleukin-1, another growth factor that lacks a signal peptide^{17,18}.

The presence of bFGF in extracts and conditioned medium from capillary endothelial cells suggested that these cells express the bFGF gene. Indeed, when poly-adenylated messenger RNA from ACE or BCE cells was hybridized with a bFGF complementary DNA probe¹³, two bFGF gene transcripts of 7.0 and 3.7 kilobases (kb) were detected in both cell types (Fig. 4, lanes *a* and *b*, respectively) but not in control cells (Fig. 4, lane *c*), whereas transcripts of the aFGF gene were detected in neither cell type (results not shown). The bFGF gene transcripts are

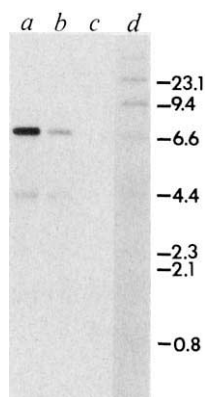


Fig. 4 Analysis of mRNA derived from ACE cells (lane *a*), BCE cells (lane *b*) or HeLa cells (controls) (lane *c*) by Northern blot and hybridization with a bFGF cDNA probe. Lanes *a-c* autoradiogram; lane *d*, size markers (kb). Capillary endothelial or HeLa cells were grown in DMEM supplemented with 10% calf serum and antibiotics⁵. Medium from subconfluent cultures was withdrawn and the cells were incubated for 3 days in the same medium containing only 0.5% CS. Cells were then lysed in 5 M guanidinium thiocyanate and polyadenylated mRNA was isolated by oligo(dT)-cellulose affinity chromatography²⁴. Aliquots of polyadenylated mRNA from the cells (5 µg per lane, each) were separated on 1.2% agarose-formaldehyde gels²⁵. The RNA was transferred to nitrocellulose filters²⁶ which were probed with a ³²P-labelled 1.4-kb *Eco*RI fragment isolated from a bovine bFGF cDNA clone¹³. Expression of bFGF gene transcripts was found in capillary endothelial cell cultures that had or had not received bFGF (1 ng ml⁻¹) (unpublished results) and throughout the time of cultivation (passages 5-16).

identical in size to those previously found in hypothalamus and the SK-Hepl cell line¹³, the correct sizes of which are 7.0 and 3.7 kb.

In conclusion, our results demonstrate that capillary endothelial cells produce and release bFGF and that bFGF derived from these cells can stimulate their proliferation in a manner similar to that described for other self-stimulating (autocrine) growth factors^{4,19}. Thus, bFGF produced by capillary endothelial cells might contribute to the formation of new capillaries.

That bFGF is present in cells as widely distributed as capillary endothelial cells suggests that it might be accessible to stimulate the proliferation of other capillary-associated FGF target cells, such as vascular smooth muscle cells, fibroblasts, chondrocytes and glial cells, and many other types². We speculate that bFGF derived from capillary endothelial cells may play an important role in embryonal growth and in the maintenance of tissue integrity. In concert with bFGF from activated macrophages²⁰, it may also participate in the tissue repair mechanisms that start upon inflammation or injury.

The identification of capillary endothelial cells as a source of bFGF might have implications for the understanding of the tumour angiogenesis process, which has previously been thought to be due to bFGF or related factors that are released from the tumours¹. However, our results suggest that the reverse phenomenon is equally conceivable: tumours might release factors able to stimulate expression, production and release of bFGF in and from capillary endothelial cells and so stimulate angiogenesis. Once the tumours are invaded by the capillaries, local release of bFGF could further enhance the growth of FGF-sensitive tumours.

We thank Dr M. Schwab for advice in preparing RNA and for reading the manuscript, Dr B. Manger for providing leukaemia cell lines, and J. Cheng for technical assistance.

Supported by grants HL 20197 and EY 02186 from the NIH, USA and by a fellowship of the Dr Mildred-Scheel Foundation to L.S.

Received 2 June; accepted 30 October 1986.

1. Folkman, J. *Cancer Res.* **46**, 467-473 (1986).
2. Gospodarowicz, D., Neufeld, G. & Schweigerer, L. *Cell Diff.* **19**, 1-17 (1986).
3. Lobb, R. *et al. J. biol. Chem.* **261**, 1924-1928 (1986).
4. Goustin, A. S., Leof, E. B., Shipley, G. D. & Moses, H. L. *Cancer Res.* **46**, 1015-1029 (1986).
5. Gospodarowicz, D., Massaglia, S., Cheng, J. & Fujii, D. *J. cell. Physiol.* **127**, 121-136 (1986).
6. Manger, B., Weiss, A., Weyand, C., Goronzy, J. & Stobo, J. *J. Immun.* **135**, 3669-3674 (1985).
7. Shing, Y. *et al. Science* **223**, 1296-1299 (1984).
8. Gospodarowicz, D., Cheng, J., Lui, G.-M., Barid, A. & Böhlen, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6963-6967 (1984).
9. Lobb, R. R. & Fett, J. W. *Biochemistry* **23**, 6925-6929 (1984).
10. Böhlen, P., Esch, F., Baird, A. & Gospodarowicz, D. *EMBO J.* **4**, 1951-1956 (1985).
11. Neufeld, G. & Gospodarowicz, D. *J. biol. Chem.* **260**, 13860-13868 (1985).
12. Gospodarowicz, D., Massaglia, S., Cheng, J., Lui, G.-M. & Böhlen, P. *J. cell. Physiol.* **122**, 323-333 (1985).
13. Abraham, J. A. *et al. Science* **233**, 545-547 (1986).
14. Moenner, M., Chevallier, B., Badet, J. & Barritault, D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5024-5028 (1986).
15. Olwin, B. B. & Hauschka, S. D. *Biochemistry* **25**, 3487-3492 (1986).
16. Gospodarowicz, D. in *Methods in Molecular and Cell Biology* (eds Barnes, D., Sirbasku, D. & Sato, G.) 275-294 (Liss, New York, 1984).
17. Auron, P. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 7907-7911 (1984).
18. March, C. J. *et al. Nature* **315**, 641-647 (1985).
19. Sporn, M. B. & Roberts, A. B. *Nature* **313**, 745-747 (1985).
20. Baird, A., Mormede, P. & Böhlen, P. *Biochem. biophys. Res. Commun.* **126**, 358-364 (1985).
21. Weber, E., Evans, C. J., Chang, J. K. & Barchas, J. J. *J. Neurochem.* **38**, 436-447 (1982).
22. Goding, J. W. *J. immunol. Meth.* **13**, 215-226 (1976).
23. Schweigerer, L. *et al. Proc. natn. Acad. Sci. U.S.A.* (submitted).
24. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
25. Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. *Biochemistry* **16**, 4743-4751 (1977).
26. Thomas, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).

A single GTP-binding protein regulates K⁺-channels coupled with dopamine, histamine and acetylcholine receptors

Kazuhiko Sasaki & Makoto Sato

Department of Physiology, School of Medicine, Iwate Medical University, 19-1, Uchimaru, Morioka, Iwate-ken, 020, Japan

Recently, a GTP-binding protein sensitive to islet activating protein (IAP) has been suggested to be important in producing K⁺-currents when the muscarinic receptor of the atrial muscle is activated by acetylcholine (ACh)¹⁻³. Here we confirm the blocking effects of IAP and GTPγS (a nonhydrolysable analogue of GTP) on the ACh-induced K⁺-current recorded from the ganglion cells of the sea slug *Aplysia* and compare their effects on histamine (HA)-induced and dopamine (DA)-induced K⁺-currents. Intracellular injections of IAP irreversibly and selectively block the openings of K⁺-channels activated by either ACh, HA, or DA without affecting the resting potential or conductance states of the membranes. Intracellular application of GTPγS alone caused extremely slow, irreversible opening of K⁺-channels; however, repetitive receptor activations significantly increase the rate of the GTPγS effect. These results strongly suggest that a GTP-binding protein such as G_i regulates the opening of K⁺-channels coupled with these receptors.

Three types of neurons were identified in the abdominal ganglion of *Aplysia*. Under voltage clamp at resting potentials, these neurons responded to ACh, HA, and/or DA with a distinct outward current⁴⁻⁶, which reversed to an inward current when the membrane was hyperpolarized below the equilibrium potential (-89 mV)⁵ for K⁺. The observed shifts of reversal potential were in very good agreement with the values predicted from the Nernst equation when the external concentration of K⁺ was altered. These responses were always associated with a marked increase in membrane conductance (Fig. 1). Thus all responses to ACh, HA, and DA were produced by an increase in membrane permeability towards K⁺.

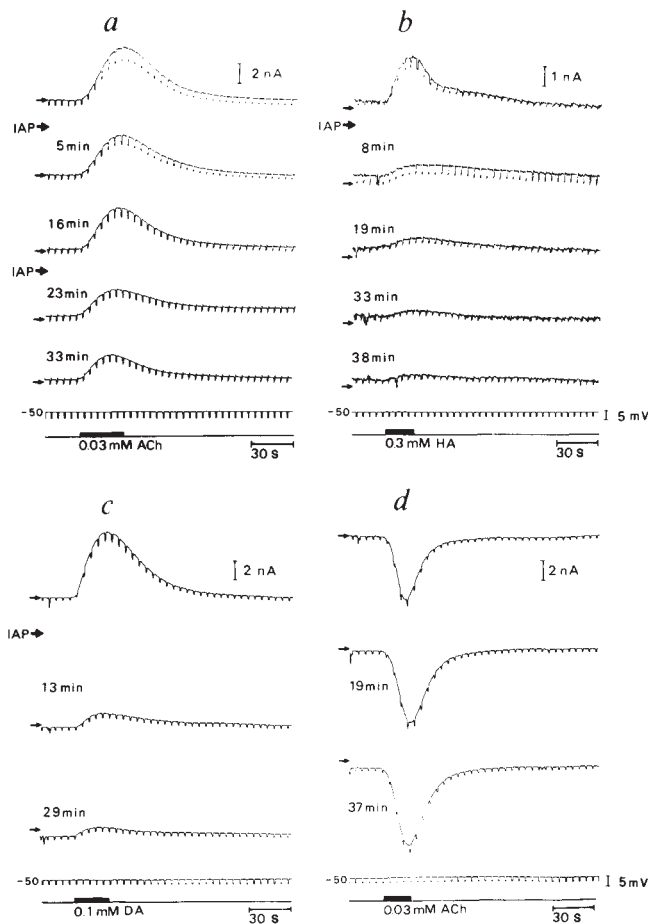


Fig. 1 Effects of IAP on the outward current responses to: *a*, ACh; *b*, HA; *c*, DA; and *d*, on the inward current response to ACh. Data shown in *a*, *b*, and *c* are from 3 different cells but *c* and *d* are from the same cell. Time after the initial intracellular injection of IAP is shown in the left of each trace. Thick arrows, times of IAP injections; thin arrows, initial level of the holding current at the clamped voltage shown (bottom). All cells were clamped at -50 mV. An upward deflection indicates an outward current. Note that the inward current response to ACh was not depressed when the outward current response of the same cell was almost completely blocked by IAP.

Methods. The abdominal ganglion of *Aplysia kurodai* was fixed in a perfusion chamber, and cells were continuously perfused with modified sea water (587 mM Na^+ , 12 mM K^+ , 671 mM Cl^- , 14 mM Ca^{2+} and 52 mM Mg^{2+} ; ref. 7). The effective perfusing volume of the chamber was 0.2 ml. Using a conventional voltage clamp method, the resting membrane potentials were clamped at -48 to -50 mV. The slope conductance of the membrane was continuously monitored by feeding repetitive (0.2 Hz) hyperpolarizing pulses (700 ms, 3 – 5 mV) to the voltage commander. IAP was dissolved in a solution including 200 mM KCl, 10 mM dithiothreitol (DTT) and 10 mM NAD $^+$. DTT was used to promote the dissociation of the A protomer from the B subunit 8 . The tip of the microelectrode was filled with a IAP solution of 5 mg ml $^{-1}$ which was applied intracellularly by a pressure pulse of 2 – 3 kg cm $^{-2}$ with a duration of 200 – 300 ms. Intracellular concentration of IAP was estimated to be 2 – 3 $\mu\text{g l}^{-1}$ by calculating a single dose of injection in the volume of the cell. ACh, DA, and HA were dissolved in the modified sea water and applied to the cells by perfusion at a flow rate of 5 ml min $^{-1}$.

In all cells examined ($n = 36$), intracellular injection of IAP caused no observable change in either the resting membrane potential or conductance, for at least one hour after injection. In contrast, all responses to ACh, HA, and DA were gradually depressed, showing almost a complete block in 40 min after an

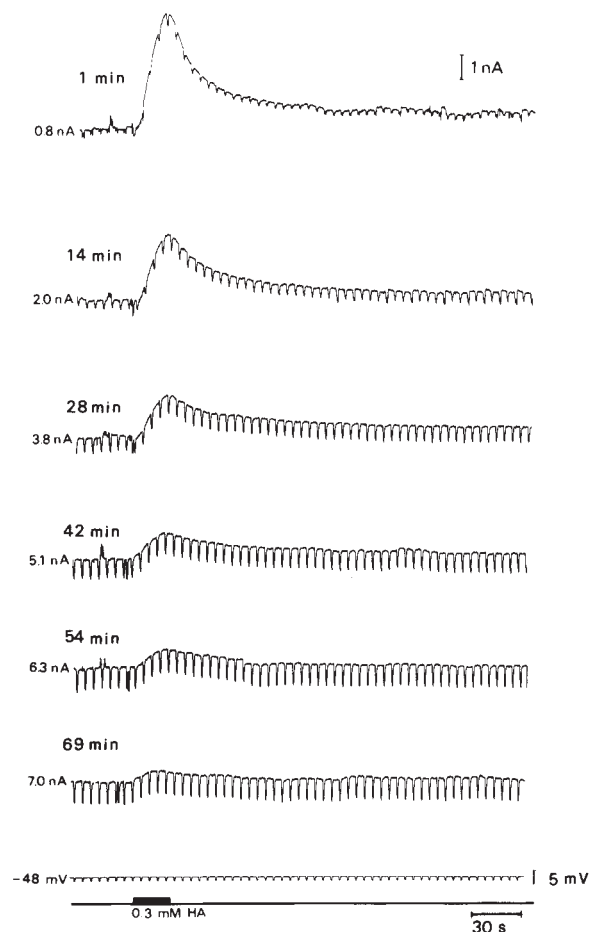
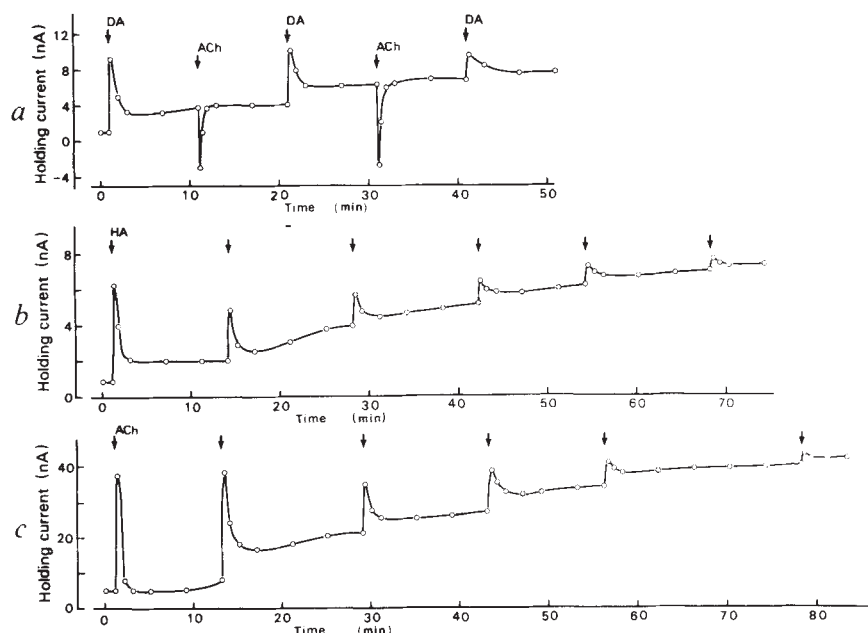


Fig. 2 Effects of GTP γ S on the HA-induced outward current responses. Intracellular application of GTP γ S was achieved by diffusion from the tip of the microelectrode filled with 0.2 M GTP γ S (Li $^+$ salt). Time after insertion of the GTP γ S electrode into the cell under study is shown on the left of each trace. The intracellular concentration of GTP γ S was ~ 50 – 100 μM at 2 – 3 h after insertion of the microelectrode. This estimation was made from separate experiments with similar electrodes filled with ^{32}P -labelled GTP. The level of the holding current (outward) at the clamped voltage of -48 mV just before testing the response is shown in nA on the left of each trace. Note the gradual increase in the holding current and the slope conductance of the membrane in contrast to the gradual decrease in the response to HA.

injection of IAP that produced an estimated intracellular concentration of 2 – 5 $\mu\text{g l}^{-1}$ (see Fig. 1). The control experiment (using an injection of IAP-solvent alone) had no effect on the responses of either receptor for a period of up to 3 h. The blocking effect of IAP was confirmed with cells having either ACh-, HA-, or DA-receptors (ACh, 15 cells; HA, 9 cells and DA, 12 cells). The effect was irreversible and very specific for the K^+ -current produced by each receptor. For example, the neuron shown in Fig. 1*c* had receptors for both DA and ACh, and responded with an outward current to DA but with an inward current to ACh (Fig. 1*d*). The ACh-induced inward current was confirmed to be Na^+ -dependent 7 . It should be noted that this response to ACh was not depressed at all by IAP when the K^+ -dependent response of the same cell to DA was completely blocked. Similar results were obtained with 7 cells having the same types of receptors. Other examples were observed in 5 cells, each of which had receptors for both γ -amino butyric acid (GABA) and DA. GABA-induced responses of these cells were all Cl^- -dependent, but DA-induced responses were K^+ -dependent. Again, IAP selectively blocked the K^+ -dependent

Fig. 3 The time courses of the changes in holding currents and responses to: *a*, DA; *b*, HA; *c*, ACh during intracellular application of GTP γ S. All cells were clamped at -50 mV throughout the experiment. The cell shown in *a* gave both outward and inward current responses to DA and ACh, respectively. Concentrations of DA, HA, and ACh were 0.1 mM, 0.3 mM, and 0.1 mM, respectively. Application time of each transmitter was 20 s (arrows). Abscissa, time after insertion of GTP γ S electrode. Note that the falling phase of the response did not return to the original resting level whenever the response was examined, and that the final level of the holding current at which no response was observable was approximately the same as the peak amplitude of the response recorded at the beginning. *a*, *b*, and *c* are representative results obtained from three different cells with their respective receptors, but similar results were obtained with 27 cells (DA, 12; HA, 5; ACh, 10).



responses without affecting the Cl^- -dependent responses, (not shown). All of these results demonstrated that the blocking action of IAP was specific to the opening of K^+ -channels coupled with these receptors. This suggested that the binding of transmitter to receptor activates the IAP-sensitive GTP-binding protein, in turn inducing opening of the K^+ -channel coupled with each receptor. Inactivation of this GTP-binding protein by IAP should result, therefore, in blocking of K^+ -channel opening.

Intracellular application of GTP γ S caused a gradual, irreversible hyperpolarization associated with a significant increase in membrane conductance. The time course of this change was extremely slow, normally taking 3 h to reach the maximal hyperpolarization of 20 – 30 mV. However, this time course was significantly speeded by frequent applications of each transmitter, as shown below. The effects of GTP γ S on the transmitter-induced responses were all analysed under the voltage clamp at -50 mV and by recording the current responses associated with the increase in slope conductance of the membranes. Accordingly, the effects of GTP γ S on the resting membranes were evaluated only by the changes in the holding current and slope conductance just before applying the transmitters. The effects of GTP γ S on the HA-induced responses are shown in Fig. 2. Similar results were obtained from 7 cells with this type of receptor. The apparent amplitude of the current response as well as the conductance response gradually decreased each time when examined repeatedly in the presence of GTP γ S. A complete block of the response was seen at 70 min. But the control experiment (single application of the transmitter at 70 min after the onset of GTP γ S) showed a much slower rate of increase in the holding current before application of the transmitter (data not shown); however, after the transmitter-induced response, there was a sudden marked increase in the holding current, which remained steady even 3 h after the addition of GTP γ S. In the presence of GTP γ S, the falling phase of the HA-induced response did not return to the original resting level, leaving a sustained residual response which was not extinguished even after complete removal of HA. Similar effects of GTP γ S were seen in the ACh-induced and DA-induced responses (Fig. 3). The residual response accumulated each time the transmitter was applied, and this fact suggested that receptor-activation promoted GDP-GTP γ S exchange at the GTP-binding site of the protein, thus speeding the effects of GTP γ S. All the K^+ -dependent, outward current responses to DA, HA, and ACh were gradually depressed during application of GTP γ S (Fig. 3).

Note that the Na^+ -dependent inward current response to ACh was rather augmented, but the K^+ -dependent outward current response to DA of the same cell was greatly depressed (Fig. 3*a*). The final level of the holding current was approximately the same as that of the peak outward current of the initial response to each transmitter. As the transmitter concentrations used in this experiment produced close to the maximum response of each cell to each transmitter, the final level of the holding current in the presence of GTP γ S might indicate a complete open state of all K^+ -channels coupled with the total number of receptors of each cell, resulting in apparent blocking of the response to each transmitter. The open state of K^+ -channels is probably maintained by irreversible binding of GTP γ S to the GTP-binding site in the α -subunit of G_i or another IAP-substrate^{9,10}, preventing the exchange reaction¹¹ between GTP and GDP that would normally close the K^+ -channels.

As shown above, IAP irreversibly blocked all K^+ -dependent responses to DA, HA, and ACh, and these transmitters in the presence of GTP γ S irreversibly opened all K^+ -channels coupled with each type of receptor. These results strongly suggested that all of these transmitter-induced increases in K^+ -conductance were produced by a GTP-dependent activation of a common mediator, such as G_i or another IAP-substrate.

Recently, the outward current responses to opiate and α_2 -agonists recorded from *locus coeruleus* neurons of the cat were shown to be blocked by IAP¹². It was also reported that the acute inhibitory effects of morphine and clonidine on the guinea pig ileum were blocked by IAP^{13,14}. The outward current responses to opiate- and α_2 -agonists and the inhibitory effects of morphine and clonidine observed could all be due to activation of K^+ -channels coupled with α_2 -receptors and morphine or opiate receptors. A number of other transmitters and hormones have been shown to cause a selective increase in K^+ -permeability of the receptor membranes in various tissues^{15–24}. It is highly possible that the activation of K^+ -channels coupled with the receptors of these transmitters and hormones is mediated by IAP-substrates^{9,10} such as N_i , and that not all of these processes are necessarily due to inhibitory effects on adenylate cyclase. Further evidence is needed before we conclude that the IAP-sensitive GTP-binding protein is directly coupled to these K^+ -channels and that its regulatory mechanism is independent of the change in cAMP, however. Also, our hypothesis appears valid only for receptor-coupled K^+ -channels that are not Ca^{2+} -dependent, since no K^+ -dependent responses observed in this

experiment were depressed in Ca^{2+} -free media or by intracellular injection of EGTA, as reported by other investigators^{6,15}.

We thank Professor M. Ui and Dr K. Nogimori for their generosity in supplying the IAP, and Professor P. Langman for his kind assistance with English usage. This research was partly supported by a Grant from the Iwate Medical University Keiryokai Research Foundation No. 26.

Received 23 June; accepted 13 November 1986.

1. Pfaffinger, P. J., Martin, J. M., Hunter, D. D., Nathanson, N. M. & Hille, B. *Nature* **317**, 536–538 (1985).
2. Breitwieser, G. E. & Szabo, G. *Nature* **317**, 538–540 (1985).
3. Sorota, S., Tsuji, Y., Tajima, T. & Pappano, A. J. *Circulation Res.* **57**, 748–758 (1985).
4. Kehoe, J. J. *J. Physiol., Lond.* **225**, 85–114 (1972).
5. Gruol, D. L. & Weinreich, D. *Brain Res.* **162**, 281–301 (1979).

6. Ascher, P. & Chesnoy-Marchais, D. *J. Physiol., Lond.* **324**, 67–92 (1982).
7. Sato, M., Austin, G., Yai, H. & Marubashi, J. *J. gen. Physiol.* **51**, 321–345 (1968).
8. Katada, T., Tamura, M. & Ui, M. *Arch. Biochem. Biophys.* **224**, 290–298 (1983).
9. Sternweis, P. C. & Robishaw, J. D. *J. biol. Chem.* **259**, 13806–13812 (1984).
10. Neer, E. J., Lok, J. M. & Wolf, L. G. *J. biol. Chem.* **259**, 14222–14229 (1984).
11. Sunyer, T., Codina, J. & Birnbaumer, L. *J. biol. Chem.* **259**, 15447–15451 (1984).
12. Aghajanian, G. K. & Wang, Y. Y. *Brain Res.* **371**, 390–394 (1986).
13. Tucker, J. F. *Br. J. Pharmacol.* **83**, 326–328 (1984).
14. Luján, M., López, E., Ramírez, R., Aguilar, H., Martínez-olmedo, M. A. & García-sainz, J. A. *Eur. J. Pharmacol.* **100**, 377–380 (1984).
15. Kehoe, J. & Marty, A. *Rev. Biophys. Bioengng.* **9**, 437–465 (1980).
16. Hartzell, H. C. *Nature* **291**, 539–544 (1981).
17. Yarowsky, P. J. & Carpenter, D. O. *Brain Res.* **144**, 75–94 (1978).
18. Gerschenfeld, H. M. & Paupardin-Tritsch, D. *J. Physiol., Lond.* **243**, 427–456 (1974).
19. Pace, C. S., Murphy, M., Conant, S. & Lacy, P. E. *Am. J. Physiol.* **233**, C164–C171 (1977).
20. Yoshimura, M. & North, R. A. *Nature* **305**, 529–530 (1983).
21. Williams, J. T., Henderson, G. & North, R. A. *Neuroscience* **14**, 95–101 (1985).
22. Cottrell, G. A., Davies, N. W. & Green, K. A. *J. Physiol., Lond.* **356**, 315–333 (1984).
23. Aghajanian, G. K. & Lakoski, J. M. *Brain Res.* **305**, 181–185 (1984).
24. Smith, P. A. *J. Physiol., Lond.* **347**, 377–395 (1984).

Recombinant interleukin-2 directly augments the cytotoxicity of human monocytes

Miroslav Malkovský, Bruce Loveland, Margaret North, Geoffrey L. Asherson, Lique Gao, Peter Ward & Walter Fiers*

MRC Clinical Research Centre, Watford Road, Harrow, Middlesex HA1 3UJ, UK

* Laboratory of Molecular Biology, State University of Ghent, Ghent, Belgium

Interleukin-2 (IL-2), originally described as a growth factor required for sustained proliferation of T cells *in vitro*^{1,2} is a glycoprotein hormone of known structure³ which appears to be important for the generation of immune responses *in vivo*⁴. As well as T lymphocytes, B lymphocytes⁵ and large granular lymphocytes with natural killer activity (NK cells)⁶ can also respond

to IL-2. The action of IL-2 seemed to be limited specifically to lymphocytes, however, and the term 'T-lymphocytotropic hormone' was used⁷. Here we provide evidence that human monocytes display a substantially increased cytotoxic activity as a direct and rapid response to human recombinant IL-2 but not to human recombinant glycosylated interferon- γ (IFN- γ) or lipopolysaccharide. Our results reveal a previously unknown function of IL-2 and suggest its possible involvement in monocyte-T cell interactions.

Rosenberg and colleagues^{8–12} have shown that lymphokine-activated killer (LAK) T cells are induced by treatment with IL-2 in the absence of antigen. These LAK effectors are non-adherent T lymphocytes⁹, which very efficiently lyse fresh autologous or allogeneic tumour cells and all cultured tumour cells tested^{8–12}. Recently it has been shown that even autologous peripheral blood mononuclear cells are susceptible to their lytic action¹³. In agreement with the data reported by Rosenberg *et al.*¹² and Sondel *et al.*¹³, we found that cytotoxic nonadherent T lymphocytes induced by IL-2 (LAK cells) could lyse all established cell lines and all other cells (including normal autologous B cells) that we used as target cells (unpublished

Table 1 Interleukin 2-induced cytotoxicity in adherent and nonadherent PBM cells

PBM cells	Lymphokine added	Expt 1 T24 targets	Expt 2a T24 targets	% Specific lysis		
				Expt 2b McSa-1 targets	Expt 3a T24 targets	Expt 3b McSa-1 targets
Nonadherent	None	3.1 (2.7)*	14.2 (3.8)	1.1 (–1.5)	4.7 (1.0)	0.2 (0.1)
Nonadherent	IL-2	91.8 (7.8)	45.9 (9.8)	52.0 (1.7)	56.4 (5.3)	54.1 (0.4)
Nonadherent	IFN- γ	3.0	17.4 (1.9)	1.2 (2.4)	7.8 (4.6)	2.3 (0)
Adherent	None	2.6 (2.7)	5.0 (5.4)	–2.1 (–0.6)	0.2 (–1.4)	0.3 (–0.1)
Adherent	IL-2	67.6 (64.0)	35.0 (33.9)	1.4 (–0.5)	68.8 (66.0)	0.1 (0.2)
Adherent	IFN- γ	4.3 (4.2)	6.1 (4.7)	–1.8 (0.3)	3.9 (2.2)	–1.2 (1.4)

Peripheral blood mononuclear cells isolated by Ficoll-Paque (Pharmacia) density gradient centrifugation were seeded (10^6 cells per well) into 96-well tissue culture plates (Nunc) in 0.2 ml of RPMI 1640 medium supplemented with 2 mM L-glutamine (Sigma), 100 IU ml^{-1} penicillin and 100 $\mu\text{g ml}^{-1}$ streptomycin (Glaxo) and 10% heat-inactivated (56 °C, 30 min) autologous or pooled allogeneic serum. After a 2-h incubation under standard conditions (at 37 °C in a humidified atmosphere of 95% air and 5% CO_2), the nonadherent cells were removed from some wells by vigorous pipetting, counted and transferred to empty wells and the volume of supplemented RPMI 1640 medium in all wells was adjusted to 0.1 ml per well. In different experiments, 15–27% of PBM cells remained strongly adherent after vigorous pipetting. In experiments 1–3, using cells isolated from 3 healthy volunteers, ~20% of cells were strongly adherent cells and 80% were nonadherent. Human IL-2 prepared by recombinant DNA techniques²⁵ and highly purified by reversed-phase high performance liquid chromatography^{14,26} (500 units per ml) or purified recombinant glycosylated human IFN- γ ^{27,28} (1,200 International Units per ml) were added to some wells and the final volume in wells was adjusted to 0.2 ml. Units of IL-2 were determined as described previously¹⁴. Endotoxin has not been detected in the purified lymphokine preparations^{26,28} using the *Limulus* amoebocyte lysate test²⁹. After a 24-h incubation under standard conditions in the presence or absence of exogenous IL-2 or IFN- γ , cell viability (89–97% in different experiments) was determined using 0.05% trypan blue and radiolabelled T24¹⁶ or McSa-1³⁰ target cells were added. The attacker to target cell ratio was 40:1 for both nonadherent and adherent effectors except for experiment 1 in which the attacker to target cell ratios were 80:1 and 20:1 for nonadherent and adherent effectors respectively. Some effector cells were treated with CAMPATH 1 and fresh autologous nonactivated serum¹⁵ before adding radiolabelled target cells and the cytotoxic activity of these lymphocyte-depleted populations is indicated in parentheses. The specific lysis of target cells was measured using a standard ⁵¹Cr-release assay³¹. In brief, after adding radiolabelled target cells, the plates were spun (400g; 5 min) and incubated for 5 h under standard conditions before re-centrifugation and removal of aliquots of supernatant for measurement of released gamma radioactivity and calculation of the per cent specific lysis. The statistical significance of data obtained in this and all subsequent sets of experiments was evaluated using Student's *t*-test. Differences between control values of lysis (cytotoxicity in the absence of tested substance) and experimental values of lysis (cytotoxicity in the presence of tested substance) were considered to be statistically significant and accepted if *P* was <0.05. Human IL-2 but not human IFN- γ induced significant cytotoxicity in all experiments.

* Values in parentheses are of effector cells treated with CAMPATH 1 monoclonal antibody and complement.

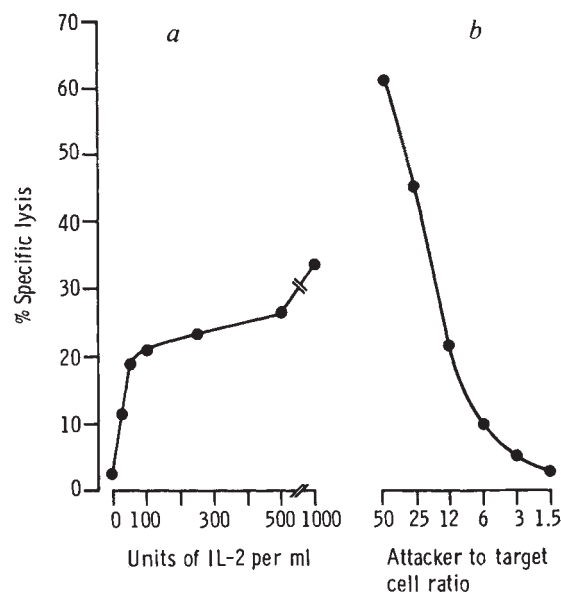


Fig. 1 The dependence of IL-2-induced monocyte cytotoxicity on the concentration of IL-2 and on the attacker to target cell ratio. Monocytes were isolated from human peripheral blood as described¹⁸ and exposed to purified recombinant IL-2 *in vitro* as described in the legend to Table 1. *a*, Cytotoxicity against T24 cells at a 20 to 1 attacker to target cell ratio after a 24-h exposure to different doses of IL-2; *b*, cytotoxicity of monocytes incubated in 500 units per ml of IL-2 at different attacker to target cell ratios. All concentrations of IL-2 (50–1,000 units per ml) induced significant cytotoxicity.

data). During these studies we noticed that human recombinant IL-2 (ref. 14) significantly enhanced the ability of cells in the strongly adherent fraction of human peripheral blood mononuclear (PBM) cells to lyse T24 cells, but not other LAK cell-sensitive targets, for example, McSa-1 cells (see Table 1). Table 1 also shows that the cytotoxic activity of LAK cells was abolished by treatment with an antilymphocyte monoclonal antibody, CAMPATH 1 (ref. 15), and autologous serum as a source of complement, confirming previously published results⁹. However, the cytotoxic activity of adherent cells induced with IL-2 was not decreased by this treatment, indicating that a nonlymphocytic population was involved. The adherent fraction of PBM cells consisted of nearly pure monocytes [up to 93% cells were immunofluorescent when stained with anti-Leu-M3 monoclonal antibody (Becton-Dickinson)]. Because IL-2 had been thought to act only on cells from the lymphocyte lineage (T, B, NK and LAK cells), we examined whether the monocyte cytotoxic activity had been directly induced by IL-2. The target cells used throughout this study were T24 cells, which were derived from a human urinary bladder transitional cell carcinoma and appeared to be relatively resistant to nonspecific NK cell killing¹⁶ but sensitive to monocyte-mediated cytotoxicity¹⁷. We first investigated the dose response and cell concentration characteristics of this phenomenon, using monocytes prepared by a published method¹⁸ (Fig. 1). The induction of cytolytic activity was directly dependent on the concentration of IL-2 (Fig. 1*a*) and, as in other cytotoxic systems, the percentage of specific ⁵¹Cr release was proportional to the attacker to target cell ratio (Fig. 1*b*). But it was possible that the few remaining lymphocytes could activate the monocytes to be cytotoxic through a monocyte cytotoxicity-augmenting factor produced as a response to IL-2. To exclude this possibility, we prepared highly purified monocytes using a solution of

Fig. 2 Effects of lipopolysaccharide (LPS), recombinant glycosylated IFN- γ , and IL-2 on the cytotoxicity of PBM cells and monocytes, with and without treatment with CAMPATH 1 monoclonal antibody plus complement and/or treatment with MHM23 monoclonal antibody. Symbols used in *a*, *b* and *c* are as follows; the percentage of rosette-forming viable cells is given by the first number in parentheses after the symbol; cell viability is given by the second number. *a*, PBM cells: $\circ$ (78%; 90%), cells only; $\bullet$ (80%; 97%), cells plus IL-2; $\blacktriangle$ (76%; 91%), cells plus IFN- γ ; ∇ (77%; 93%), cells plus LPS. *b*, Monocytes: $\circ$ (2%; 94%), cells only; $\bullet$ (3%; 93%), cells plus IL-2; $\blacktriangle$ (2%; 95%), cells plus IFN- γ ; ∇ (2%; 96%), cells plus LPS. *c*, Monocytes pretreated with CAMPATH 1 and autologous serum: $\circ$ (<1%; 92%), cells only; $\bullet$ (<1%; 92%), cells plus IL-2. $\square$ (all panels), cytotoxic assays performed in the presence of MHM23 monoclonal antibody (cells only); $\blacksquare$, with MHM23 (cells plus IL-2). **Methods.** PBM cells or monocytes were isolated from human peripheral blood using Ficoll-Paque (Pharmacia) or Nycodenz Monocytes (Nyegaard) respectively, according to the instructions of the manufacturers. The resulting cell populations were exposed for 24 h to either purified recombinant IL-2 (500 units per ml) or purified recombinant glycosylated IFN- γ (1,200 International Units per ml) as described in the legend to Table 1 or 50 μ g ml⁻¹ of lipopolysaccharide from *Escherichia coli* 0127:B8 (Difco). To deplete the monocyte population of residual contaminating lymphocytes before the exposure to IL-2, a part of the monocyte preparation was treated with the rat monoclonal antibody, CAMPATH 1, and autologous noninactivated serum, as described¹⁵. The purity of monocytes determined by peroxidase stain was >99%. After all treatments, the specific lysis of radiolabelled T24 cells was determined as described in the legend to Table 1. MHM23 (refs 19, 20) ascites fluid was used at a final dilution of 1 in 200 to block lysis and was added to the cells at the time of mixing effectors with targets. In parallel samples, the number of cells forming rosettes with sheep erythrocytes was measured. For rosette assays, mononuclear cells (2.5×10^5 cells in 50 μ l) were mixed with 50 μ l of washed (in phosphate-buffered saline, PBS) sheep erythrocytes (1% suspension in PBS) and 50 μ l of heat-inactivated foetal calf serum, briefly centrifuged (400g, 5 min, 4°C), and incubated overnight at 4°C. The pellet was gently resuspended in 50 μ l of supernatant and was mixed with 50 μ l of PBS containing 50 μ g ml⁻¹ fluorescein diacetate. Viable rosette-forming cells were counted using a Zeiss fluorescence microscope. Samples of some cell suspensions were assayed by immunofluorescent staining before the cytotoxic assay. The monoclonal antibodies used were anti-Leu-11, anti-Leu-M3, anti-Leu-4 (Becton Dickinson) and CAMPATH 1. The percent positive staining, in respective order, for each cell suspension was: untreated PBM cells (6, 21, 66 and 82), IL-2-treated PBM cells (5, 19, 69 and 82), untreated monocytes (<1, 92, <1 and 3), IL-2-treated monocytes (<1, 92, <1 and 3), and IL-2-treated monocytes pretreated with CAMPATH 1 plus autologous serum (<1, 93, <1 and <1). The statistical analysis using Student's *t*-test revealed that IL-2, but not IFN- γ or LPS, significantly augmented cytotoxicity. Moreover, a statistically significant inhibition of PBM cell-mediated cytotoxicity, but not monocyte-mediated cytotoxicity, occurred in the presence of MHM23 monoclonal antibody.

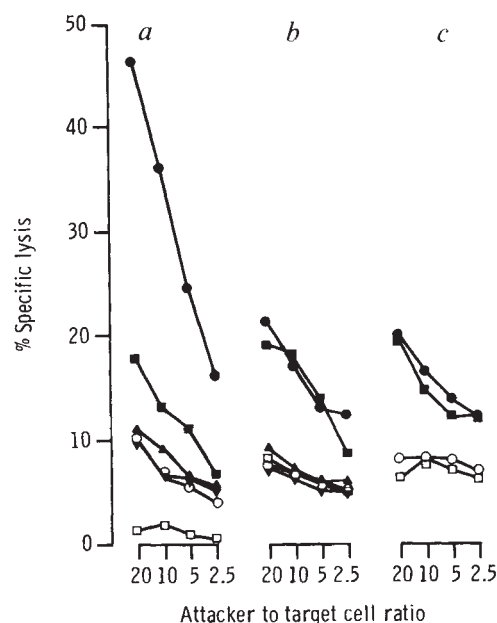


Fig. 3 Reduced cytotoxic activity of effector monocytes following a monocyte-specific lytic treatment. *a*, PBM cells; *b*, monocytes; ○, without IL-2; ●, with IL-2; ■, incubation in medium with IL-2 followed by treatment with anti-Leu-M3 and anti-mouse Ig antibodies plus complement; □, incubation in medium with IL-2 and treatment with anti-Leu-M3 plus complement; △, incubation in medium with IL-2 and treatment with anti-mouse Ig plus complement; ▲, incubation in medium with IL-2 and treatment with anti-Leu-M3 and anti-mouse immunoglobulin (controls for the specificity of complement-mediated lysis).

Methods. PBM cells and monocytes were isolated from human peripheral blood and the monocyte preparation was depleted of residual contaminating lymphocytes by the CAMPATH 1 treatment as described (Fig. 2 legend). Unseparated PBM cells and purified monocytes were exposed for 24 h to recombinant IL-2 (500 units per ml). The cells were then treated with anti-Leu-M3 (a 1:10 dilution, 40 min, 4°C), washed, treated for 40 min at 4°C with a 1:10 dilution of rabbit anti-mouse Ig polyclonal antibody (Nordic), washed, and treated with a 1:20 dilution of rabbit complement absorbed with human leucocytes. After all treatments, the specific lysis of radiolabelled T24 cells was determined as described in the legend to Table 1. Before the cytotoxic assay, the IL-2-treated cell preparations were examined by immunofluorescence with anti-Leu-M3 and CAMPATH 1. The percentages of Leu-M3-positive cells and CAMPATH 1-positive cells were 20% and 79% for PBM cells and 92% and <1% for purified monocytes, respectively. The monocyte-specific treatment with antibodies and complement lysed 17% of unseparated PBM cells and 70% of monocytes as judged by Trypan blue staining and caused a statistically significant inhibition of monocyte-mediated cytotoxicity.

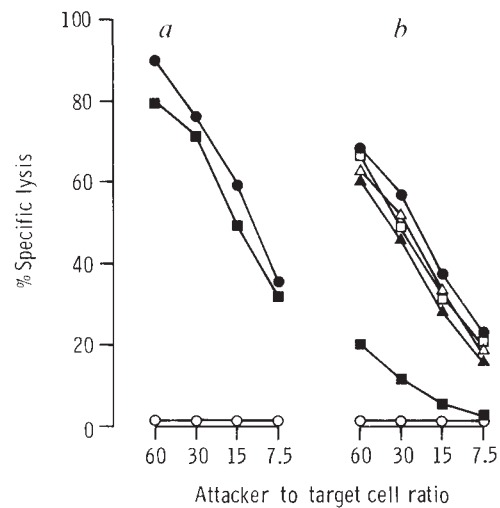
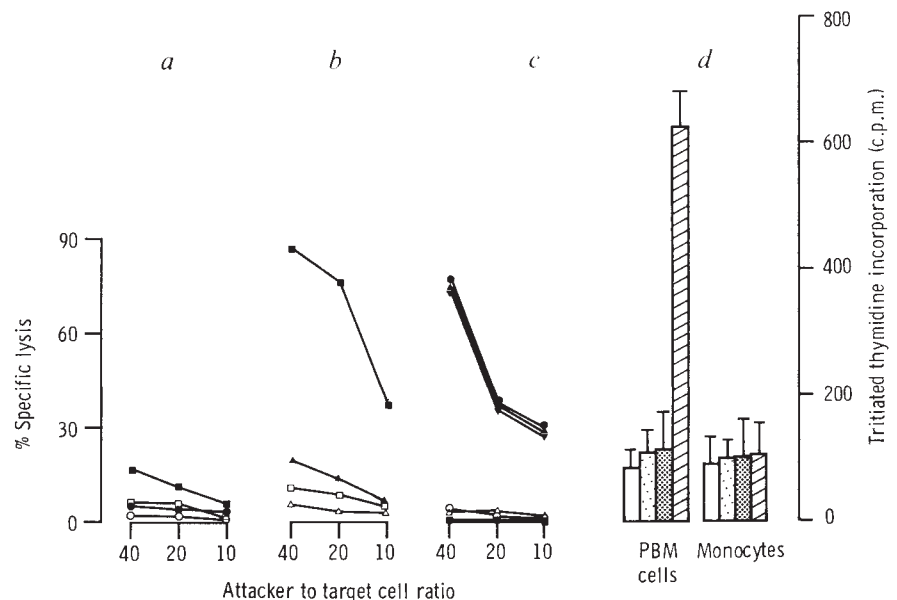


Fig. 4 Induction of monocyte cytotoxic activity by IL-2; proliferation assay and test for mediation by soluble factor. *a*, ■, fresh PBM cells with IL-2; □, PBM cells with no IL-2; ●, fresh monocytes with IL-2; ○, fresh monocytes, no IL-2 (note that the very brief exposure to IL-2 was sufficient to increase the cytotoxic potential of cells, but IL-2 itself had no effect on the chromium release from radiolabelled T24 cells—data not shown). *b*, PBM cells: ▲, treated with IL-2 and CAMPATH 1 + complement; ■, with IL-2, no CAMPATH 1; △, no IL-2, with CAMPATH 1 and complement; □, no IL-2, no CAMPATH 1. *c*, Monocytes: ○, incubated in medium without IL-2; ●, incubated with IL-2; △, treated with CAMPATH 1 after incubation without IL-2; ▲, treated with IL-2 and CAMPATH 1; ▼, treated with CAMPATH 1 before and after incubation in medium with IL-2, and MHM23 present during the cytotoxic assay; ■, T24 target cells exposed to supernatants from monocytes treated with IL-2. *d*, Proliferation assay: □, fresh cells without IL-2; ▤, fresh cells with IL-2; ▨, cells incubated 24 h without IL-2; ▩, cells incubated 24 h with IL-2.

Methods. PBM cells and monocytes were

obtained from human peripheral blood and exposed to purified recombinant IL-2 (500 units per ml) as described in the legend to Fig. 2. Some samples were treated with CAMPATH 1 and nonactivated autologous serum¹⁵ after the exposure to IL-2. One monocyte sample received the same treatment with CAMPATH 1 before as well as after the exposure to IL-2 and, in addition, MHM23 monoclonal antibody^{19,20} (a final concentration of 1 in 200 ascites) was added just before incubation with radiolabelled T24 cells. The cytotoxicity was assessed following the 24-h incubation with or without IL-2 (*b* and *c*) as well as immediately after the cells were obtained from separation gradients (*a*), also, with or without IL-2 (500 units per ml) present in the assay. Supernatants were taken after 24 h from 0.2 ml cultures of monocytes incubated at concentrations equivalent to attacker to target ratios of 40:1, 20:1 and 10:1, with IL-2 and assessed for the presence of cytotoxic factors by addition to T24 target cells. DNA synthesis (panel *d*) was measured³² in both freshly isolated cells and after 24 h in culture. Briefly, 10^5 cells were pulsed with $0.5 \mu\text{Ci } ^3\text{H}$ -thymidine for 5 h and subsequently the incorporation of tritiated thymidine was measured by liquid scintillation counting as described previously³². The statistical analysis (Student's *t*-test) of control and experimental values indicated that (1) the IL-2-induced cytotoxicity was significant; (2) the treatment with CAMPATH 1 plus complement significantly reduced the cytotoxic activity of PBM cells; (3) the monocyte-mediated cytotoxicity was not significantly influenced by the treatment with CAMPATH 1 plus complement both before and after the IL-2 induction and by the presence of MHM23 monoclonal antibody; (4) the supernatant from monocytes treated with IL-2 was not significantly cytotoxic for T24 cells; (5) IL-2 induced a significant incorporation of ^3H -thymidine into PBM cells, but not into monocytes.



Nycodenz Monocytes (Nyegaard). The monocytes were treated with CAMPATH 1 monoclonal antibody¹⁵ and autologous serum to deplete any residual lymphocytes before exposure to IL-2. In addition, during the cytotoxicity assay, some wells contained MHM23 monoclonal antibody which reacts with the β p95 chain of the LFA-1 molecule and is known to inhibit cytotoxic T cells and NK cells^{19,20} (Fig. 2). Neither the treatment with CAMPATH 1 monoclonal antibody, nor the presence of MHM23 monoclonal antibody, nor the combination of both influenced the cytotoxic activity of monocytes. However, the lytic function of unseparated PBM cells was severely diminished by MHM23 monoclonal antibody. These results indicated that the cytotoxic cells were not T lymphocytes or NK cells. Also, neither lipopolysaccharide, nor purified recombinant glycosylated human IFN- γ augmented the cytotoxicity of PBM cells (confirming previous results²¹) or of monocytes (Table 1 and Fig. 2). However, the same preparation of IFN- γ enhanced the expression of major histocompatibility complex antigens on human monocytes (ref. 22 and J. Farrant, personal communication). It is noteworthy that although the dose and batch of lipopolysaccharide we used did not induce any significant monocyte cytotoxicity after a 24-h exposure (8% specific ⁵¹Cr release), a very significant enhancement of cytotoxic activity of human monocytes was measured after a 48-h incubation period (57% specific ⁵¹Cr release). Cytolytic activity was greatly reduced after monocytes were treated with anti-Leu-M3 (complement nonfixing, first layer) and rabbit anti-mouse Ig polyclonal antibody (second layer) plus rabbit complement (Fig. 3). This treatment lysed 70% of viable monocytes and depleted the cytolytic activity to a similar extent. As previously reported, the cytolytic effect of monocytes against T24 cells did not appear to be mediated directly by a cytotoxic soluble factor¹⁷ or to require DNA synthesis for its induction (Fig. 4). Furthermore, the possibility that an IL-2-induced, PBM cell-secreted factor enhanced the cytotoxicity of monocytes seems unlikely, since the supernatant from PBM cells pulsed with IL-2 had no effect on monocyte cytotoxicity (data not shown). On the basis of these findings, we conclude that monocyte function can be directly influenced by IL-2. Moreover, our data are compatible with a report²³ showing that normal human monocytes express receptors for IL-2 on their surface. These observations, together with a recent study²⁴, show that IL-2 is not exclusively a lymphocyto-trophic hormone. The major physiological role of IL-2 was thought to be activation of resting T lymphocytes (in G₁) to progress through the proliferative phases (S, G₂ and M) of the cell cycle. The biological significance of the effect of IL-2 on monocytes has yet to be determined, but our results imply that the IL-2 system may influence the behaviour of monocytes and their interactions with other cells during an immune response.

We thank Professor A. J. McMichael for the gift of the MHM23 monoclonal antibody, Dr C. D. L. Reid for providing CAMPATH 1, Dr Nancy M. Hogg for helpful discussions, Mrs Margaret Runnicles for typing the manuscript and Mrs Christine Gilderdale for drawing the figures. We also thank Biogen, S.A. for their support and supplying us with IL-2.

Received 14 August; accepted 23 October 1986.

1. Morgan, D. A., Ruscetti, F. M. & Gallo, R. C. *Science* **193**, 1007-1008 (1976).
2. Gillis, S. & Smith, K. A. *Nature* **268**, 154-156 (1977).
3. Robb, R. J. *Immun. Today* **5**, 203-209 (1984).
4. Malkovsky, M. & Medawar, P. B. *Immun. Today* **5**, 340-343 (1984).
5. Waldmann, T. A. *et al. J. exp. Med.* **160**, 1450-1466 (1984).
6. Ortaldo, J. R. *et al. J. Immun.* **133**, 779-783 (1984).
7. Smith, K. A. *Rev. Immun.* **2**, 319-333 (1984).
8. Yron, I., Wood, T. A., Spiess, P. J. & Rosenberg, S. A. *J. Immun.* **125**, 238-245 (1980).
9. Grimm, E. A., Mazumder, A., Zhang, H. Z. & Rosenberg, S. A. *J. exp. Med.* **155**, 1823-1841 (1982).
10. Grimm, E. A. *et al. J. exp. Med.* **157**, 884-897 (1983).
11. Grimm, E. A. *et al. J. exp. Med.* **158**, 1356-1361 (1983).
12. Rosenberg, S. A. *et al. Science* **223**, 1412-1415 (1984).
13. Sondel, P. M. *et al. J. Immun.* **137**, 502-511 (1986).
14. Malkovsky, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 536-538 (1985).
15. Hale, G. *et al. Blood* **62**, 873-882 (1983).

16. Bubenik, J. *et al. Neoplasia* **24**, 319-326 (1978).
17. Rinehart, J. J., Lange, P., Gormus, B. J. & Kaplan, M. E. *Blood* **52**, 211-220 (1978).
18. Zanella, A. *et al. J. immun. Meth.* **41**, 279-288 (1981).
19. Hildreth, J. E. K., Gotch, F. M., Hildreth, P. D. K. & McMichael, A. J. *Eur. J. Immun.* **13**, 202-208 (1983).
20. Dongworth, D. W., Gotch, F. M., Hildreth, J. E. K., Morris, A. & McMichael, A. J. *Eur. J. Immun.* **15**, 888-892 (1985).
21. Devos, R., Plaetinck, G. & Fiers, W. *Eur. J. Immun.* **14**, 1057-1060 (1984).
22. Kelley, V. E., Fiers, W. & Strom, T. B. *J. Immun.* **132**, 240-245 (1984).
23. Herrmann, F., Cannistra, S. A., Levine, H. & Griffin, J. D. *J. exp. Med.* **162**, 1111-1116 (1985).
24. Benveniste, E. N. & Merrill, J. E. *Nature* **321**, 610-613 (1986).
25. Devos, R. *et al. Nucleic Acids Res.* **11**, 4307-4323 (1983).
26. Liang, S.-M. *et al. Biochem. J.* **229**, 429-439 (1985).
27. Scathill, S. J., Devos, R., Van Der Heyden, J. & Fiers, W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4654-4658 (1983).
28. Devos, R., Opsomer, C., Scathill, S. J., Van Der Heyden, J. & Fiers, W. *J. Interferon Res.* **4**, 461-468 (1984).
29. Levin, J., Tomasulo, P. A. & Oser, R. S. *J. Lab. clin. Med.* **75**, 903-911 (1970).
30. Malkovsky, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 6322-6326 (1983).
31. Malkovsky, M., Asherson, G. L., Stockinger, B. & Watkins, M. C. *Nature* **300**, 652-655 (1982).
32. Malkovsky, M. *et al. J. Immun.* **130**, 785-790 (1983).

Tumour necrosis factor and lymphotoxin genes map close to H-2D in the mouse major histocompatibility complex

Urs Müller*, C. Victor Jongeneel†, Sergei A. Nedospasov‡, Kirsten Fischer Lindahl§ & Michael Steinmetz*||

* Basel Institute for Immunology, Grenzacherstrasse 487, CH-4058 Basel, Switzerland

† Ludwig Institute for Cancer Research, Lausanne Branch, CH-1066 Epalinges, Switzerland ‡ Institute of Molecular Biology, USSR Academy of Sciences, Moscow, USSR

§ Howard Hughes Medical Institute, Departments of Microbiology and Biochemistry, University of Texas Health Science Center at Dallas, Dallas, TX 75235-9050, USA

Tumour necrosis factor (TNF- α) and lymphotoxin (TNF- β) are related proteins, secreted by macrophages and lymphocytes respectively, which play a role in destruction of tumour cells and virally infected cells (for reviews see refs 1, 2). TNF- α is a non-glycosylated protein of relative molecular mass 17,000 (*M_r* 17 K), whereas TNF- β is a glycoprotein of *M_r* 25 K. Both TNF- α and TNF- β aggregate into multimers and act through the same receptor molecule on target cells. Genes encoding these two TNF proteins have been cloned from mouse and man³⁻⁶ and in both are closely linked, being separated by ~1 kilobase (kb) of DNA^{7,8}. In the mouse these genes are located on chromosome 17 (ref. 8), but in man they are on the short arm of chromosome 6 (ref. 9). This segment of chromosome 6 also contains the genes of the major histocompatibility complex (MHC), as does chromosome 17 in the mouse. To find out whether the TNF genes are located within the MHC, we used polymorphic restriction sites to analyse a panel of MHC congenic and intra-MHC recombinant mouse strains. Initially, we mapped the TNF genes the *D* or *Qa* region in the distal half of the mouse MHC. We then studied a gene cluster encompassing part of the *D* and *Qa* regions and found the TNF genes are located 70 kb proximal to the *D* gene.

A portion of exon 4 of the TNF- α gene was used as a molecular probe to search for polymorphic restriction sites in different inbred mouse strains. Mouse strains AKR, A.SW, CAS3, C3H/HeJ, and C57BL/10 contain a *Bam*HI fragment of ~12 kb, and strains A/J, BALB/c and DBA/2 possess a smaller *Bam*HI fragment of ~10 kb. To confirm that the TNF genes are linked to the MHC, we analysed strains B10.A and B10.D2. The MHC in these mice is derived from strains A/J and DBA/2, respectively, but has been placed on the genetic background of

|| Present address: Central Research Units, F. Hoffmann-La Roche & Co., Ltd. CH-4002 Basel, Switzerland.

Table 1 MHC alleles of inbred and recombinant mouse strains

Strain	MHC donor strains	Genetic locus							
		<i>K</i>	<i>I</i>	<i>S</i>	<i>D</i>	<i>Qa-2</i>	<i>Tla</i>	<i>Qa-1</i>	<i>Hmt</i>
A/J		k	k	d	d	a	a	a	a
AKR		k	k	k	k	b	b	b	a
A.SW		s	s	s	s	a	b	b	a
BALB/c		d	d	d	d	a	c	b	a
C3H/HeJ		k	k	k	k	b	b	b	a
C57BL/10		b	b	b	b	a	b	b	a
CAS3		c3	c3	c3	w3	a	c3	c3	b
DBA/2		d	d	d	d	a	c	b	a
B10.A	A/WySnSg	k	k	d	d	a	a	a	a
B10.D2	DBA/2	d	d	d	d	a	c	b	a
C3H.OH	DBA/2, C3H/J	d	d	d	k	b	b	b	a
C3H.OL	DBA/2, C3H/J	d	d	k	k	b	b	b	a
CAS3(R1)	CAS3, C3H/HeJ	c3	c3	c3	w3	a	b	b	a
CAS3(R4)	C3H/HeJ, CAS3	k	k	k	k	b	b	b	b
CAS3(R11)	C3H.SW, CAS3, C3H/HeJ	b	c3	c3	k	b	b	b	a

Alleles, based on serological identification, are listed according to refs 10 and 11. Locations of recorded crossovers are indicated by vertical lines.

C57BL/10. The smaller *Bam*HI fragment was found in both these congenic strains (Fig. 1a), confirming the linkage of the TNF genes to the MHC.

Next we analysed two congenic mouse strains containing a recombinant MHC derived from two haplotypes distinguishable for the TNF genetic marker. In mouse strains C3H.OH and C3H.OL, the proximal portion of the MHC (oriented towards the centromere) is of DBA/2 origin, whereas the distal portion (oriented towards the telomere) is of C3H origin (see Table 1 and ref. 10). The recombination occurred between *I* and *S* in strain C3H.OL, and between the *S* and *D* region marker loci in C3H.OH. Southern blot analysis of both strains using the TNF- α probe revealed the larger *Bam*HI fragment as in C3H mice and therefore located the TNF genes distal to the *S* region (Figs 1a, 2a).

We then analysed three intra-MHC recombinant mouse strains, CAS3(R1), CAS3(R4) and CAS3(R11), obtained from matings between C3H/HeJ and CAS3 mice¹¹. We found a *Bgl*II fragment of ~23 kb with the TNF probe in CAS3 and a smaller fragment of ~19 kb in C3H/HeJ (Fig. 1b). Analysis of the three recombinant strains (Fig. 1b) revealed the larger fragment in CAS3(R1) and the smaller one in CAS3(R4) and CAS3(R11). Comparison of this result with the location of the breakpoints and the constellation of the recombinant haplotypes maps the TNF genes distal to the *S* and proximal to the *Tla* region marker loci, within the *D* or *Qa* region of the mouse MHC (Fig. 2a).

A 500-kb gene cluster with 13 class I genes, spanning the breakpoint between the *D* and *Qa* regions, has been cloned from the MHC of the BALB/c mouse¹² (Fig. 2b). To find out whether this cluster also contains the TNF genes, we analysed cosmid clones constituting the cloned region¹² by dot-blot hybridization. Only cosmid clone II 3.5, located at the proximal end of the cluster¹², gave a positive signal as strong as that obtained from the TNF- α gene used as a positive control (data not shown). This clone was digested with several restriction enzymes and analysed by Southern blot hybridization (Fig. 3a). The results, together with a comparison of the restriction maps of cosmid clone II 3.5 (ref. 12) and a mouse TNF genomic clone⁸, located the TNF- α and - β genes to a region ~70 kb upstream of the *D* gene (Figs 2b and 3b). The TNF genes are transcribed in the opposite direction to the class I genes of this cluster.

The MHC, a large genetic region on mouse chromosome 17, encodes class I and class II cell-surface molecules that present foreign antigens to T lymphocytes (for reviews see refs 13, 14). Several class I and class II genes with unknown functions and 6 genes encoding complement components and steroid 21-

hydroxylase have been identified in the mouse MHC (Fig. 2b). The location of the TNF genes at the proximal end of the *D/Qa* region gene cluster is close to genes in the MHC which are also structurally and functionally unrelated to class I and class II genes. It is curious that all these genes (C2/Bf; Slp/C4; 21-OHA/21-OHB; and TNF- α /TNF- β) are organized in pairs of related genes. It is not known whether the tight linkage between genes encoding molecules involved in the recognition (class I and class II) and effector (complement and TNF) phases of the immune response has any functional significance. TNF- α augments transcription of class I genes¹⁵.

In man, certain diseases are associated with the presence of particular MHC alleles in affected individuals (for review see

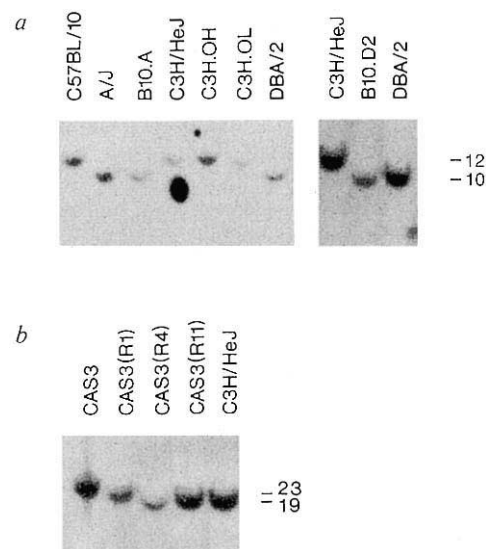
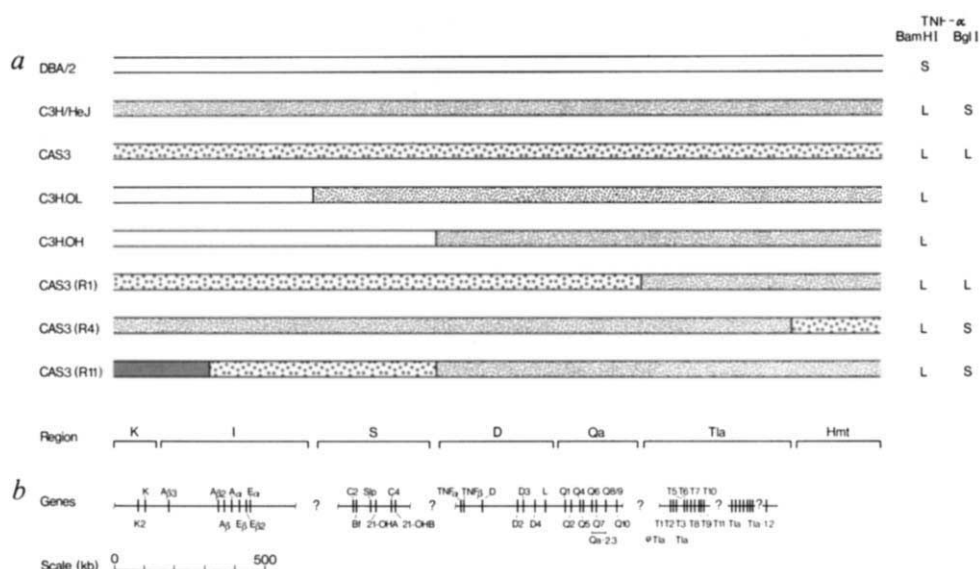


Fig. 1 Southern blot analysis of polymorphic restriction fragments containing the TNF- α gene. *Bam*HI (a) and *Bgl*II (b) were used to digest liver DNA from the mouse strains indicated. After agarose gel electrophoresis and transfer to Zeta probe membranes²⁰, hybridization was as described²¹ using as a probe an oligolabelled²² 450-bp *Hind*III-*Eco*RI fragment derived from exon 4 of the TNF- α gene⁸. Sizes of the TNF-specific restriction fragments (in kb) were determined using restriction fragments of phage λ as markers.

Fig. 2 Location of the TNF- α and - β genes in the mouse MHC. *a*, Polymorphic *Bam*HI and *Bgl*II fragments map the closely linked TNF- α and - β genes to the *D* or *Qa* region of the MHC. Different MHC haplotypes are distinguished by different textures to highlight the origin of MHC DNA in the recombinant mouse strains. Vertical lines, recombinational breakpoints; letters, regions of the MHC defined by them. *Hmt*, a gene which interacts with a maternally inherited factor to determine a cell-surface antigen (Mta), maps between the *Tla* and *Upg*-1 marker loci, about 2 cM distal to *D* (ref. 23) and is separated from *Tla* by the recombinational breakpoint in mouse strain CAS3(R4) (ref. 15). TNF- α alleles, as determined by Southern blot analysis using the restriction enzymes *Bam*HI and *Bgl*II (Fig. 1), are shown as S (small fragment) and L (large fragment). *b*, Gene clusters cloned from the MHC of the BALB/c mouse drawn to scale and identified genes are indicated



by vertical lines (additional data from refs 12, 21, 24, 25). Class I genes map to the *K*, *D*, *Qa* and *Tla* regions, class II genes to the *I* region and the complement and 21-hydroxylase genes to the *S* region. ?, Unknown molecular distance between gene clusters. All class I genes of the *Tla* region are located between the recombinational breakpoints in CAS3(R1) and CAS3(R4) (M. Kiefer, K.F.L. and M.S., unpublished data).

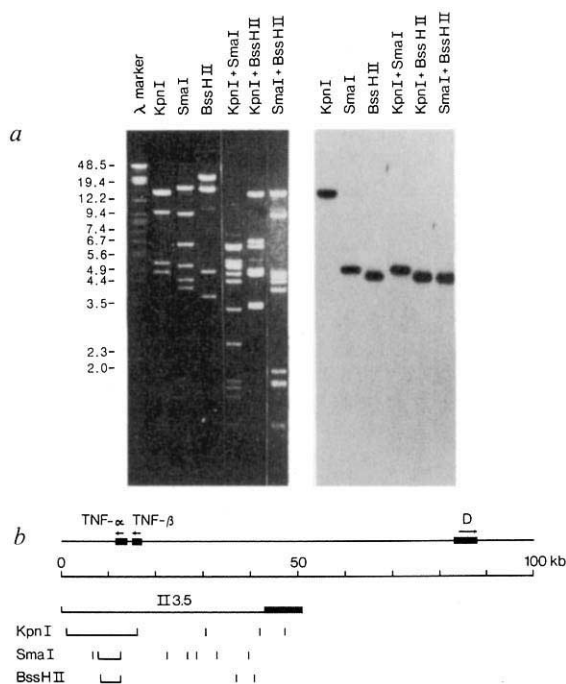


Fig. 3 Location of the TNF genes 70 kb proximal to the *D* gene in the BALB/c mouse. *a*, Fragments resulting from single and double digestions of cosmid clone II 3.5 using several enzymes (*Kpn*I, *Sma*I and *Bss*HII are shown) were separated by agarose gel electrophoresis (left panel), transferred to a Zeta probe membrane and hybridized with the TNF- α probe (right panel) as in Fig. 1. Sizes of phage λ marker fragments are given in kb. *b*, Location and orientation of the TNF genes on cosmid clone II 3.5 based on hybridization analysis and restriction map comparison. *Kpn*I, *Sma*I and *Bss*HII restriction sites are shown for the region defined by clone II 3.5 only¹². Solid box, cosmid vector portion of clone II 3.5; horizontal lines, restriction fragments hybridizing to the probe and therefore containing exon 4 of TNF- α . Orientation of transcription (5' $\rightarrow$ 3') is indicated by arrows. For a complete map of the 500-kb gene cluster see Fig. 2*b* and ref. 12.

ref. 16). From their location in the mouse MHC, the TNF genes are presumably closely linked to the HLA-B gene in man, a location consistent with recent mapping data¹⁷. Several rheumatic diseases are associated with the HLA-B27 allele, notably ankylosing spondylitis, an inflammatory disease of vertebral joints¹⁶. It is still unknown whether ankylosing spondylitis is caused by the B27 allele itself or by an allele of a second gene in linkage disequilibrium with B27. TNF- α stimulates collagenase production by human synovial cells¹⁸ and resorption of proteoglycan in cartilage¹⁹. Thus an abnormal TNF gene could indeed be involved in ankylosing spondylitis and other inflammatory diseases.

We thank S. Carson and H. von V. Boehmer for comments on the manuscript and Ms Judie Hossmann for typing. The Basel Institute for Immunology was founded and is supported by F. Hoffmann-La Roche, Ltd, Basel, Switzerland.

Received 19 September; accepted 19 November 1986.

- Beutler, B. & Cerami, A. *Nature* **320**, 584-588 (1986).
- Old, L. J. *Science* **230**, 630-632 (1985).
- Gray, P. W. *et al.* *Nature* **312**, 721-724 (1984).
- Pennica, D. *et al.* *Nature* **312**, 724-729 (1984).
- Fransen, L. *et al.* *Nucleic Acids Res.* **13**, 4417-4429 (1985).
- Pennica, D., Hayflick, J. B., Bringman, T. S., Palladino, M. A. and Goeddel, D. V. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6060-6064 (1985).
- Nedospasov, S. A. *et al.* *Dokl. Akad. Nauk SSSR* **285**, 1487-1490 (1985).
- Nedospasov, S. A. *et al.* *Nucleic Acids Res.* **14**, 7713-7727 (1986).
- Nedwin, G. W. *et al.* *Nucleic Acids Res.* **13**, 6361-6373 (1985).
- Klein, J., Figueroa, F. & David, C. S. *Immunogenetics* **17**, 553-596 (1983).
- Fischer Lindahl, K. *Curr. Top. Microbiol. Immun.* **127**, 272-278 (1986).
- Stephan, D. *et al.* *J. exp. Med.* **163**, 1227-1244 (1986).
- Flavell, R. A. *et al.* *Science* **233**, 437-443 (1986).
- Steinmetz, M. in *Genetic Engineering* Vol. 5 (ed. Rigby, P. W. J.) 117-158 (Academic, London 1986).
- Collins, T., Lapiere, L. A., Fiers, W., Strominger, J. L. & Pober, J. S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 446-450 (1986).
- Thomson, G. in *Handbook of Experimental Immunology* Vol. III (ed. Weir, D. M.) 102.1-102.12 (Blackwell, Oxford, 1986).
- Spies, T. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 8699-8702 (1986).
- Dayer, J.-M., Beutler, B. & Cerami, A. *J. exp. Med.* **162**, 2163-2168 (1985).
- Saklatvala, J. *Nature* **322**, 547-549 (1986).
- Reed, K. C. and Mann, D. A. *Nucleic Acids Res.* **13**, 7201-7221 (1985).
- Steinmetz, M., Stephan, D. & Fischer Lindahl, K. *Cell* **44**, 895-904 (1986).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266-267 (1984).
- Fischer Lindahl, K., Hausmann, B. & Chapman, V. M. *Nature* **306**, 383-385 (1983).
- Steinmetz, M., Winoto, A., Minard, K. & Hood, L. *Cell* **28**, 489-498 (1982).
- Steinmetz, M. *et al.* *EMBO J.* **3**, 2995-3003 (1984).

A transcription factor which binds to the enhancers of SV40, immunoglobulin heavy chain and U2 snRNA genes

Dirk Bohmann, Walter Keller, Trevor Dale*, Hans R. Schöler*, Graham Tebb† & Iain W. Mattaj†

Division of Molecular Biology, Institute of Cell and Tumor Biology, German Cancer Research Center, D-6900 Heidelberg, FRG
† European Molecular Biology Laboratory (EMBL), D-6900 Heidelberg, FRG.

In eukaryotes the transcriptional control of RNA polymerase II-mediated gene expression is exerted by *cis*-acting regulatory DNA elements classified as promoter and enhancer sequences. These elements are composed of a number of different protein binding sites (see refs 1, 2 for review). The regulatory factors that recognize such 'modules' may be ubiquitous, tissue- or stage-specific, and positively or negatively acting. According to this model the transcriptional activity of a given gene is programmed by a combination of different modules. We analysed such a site of protein-DNA interaction, the octamer motif³⁻⁶, in the enhancers of the simian virus (SV40) early genes and the murine immunoglobulin heavy-chain gene, and in the distal sequence element (DSE) of the U2 small nuclear (sn)RNA gene of *Xenopus laevis*. The corresponding DNA-binding factor appears to be the same in the three cases. Moreover, a fraction containing partially purified octamer motif binding factor has a stimulatory effect on transcription in an *in vitro* system.

We and others have previously demonstrated stimulation of transcription by the SV40 enhancer in an *in vitro* system derived from cultured cells⁷⁻¹⁰. Accordingly, transcription extracts should contain factors that can bind to enhancer and promoter sequences. To monitor this binding directly, we performed DNase I-footprinting experiments¹¹ on the SV40 enhancer region (Fig. 1a, SV40). Three areas of nuclease protection flanked by DNase-hypersensitive sites could be distinguished in each of the two 72-base pair (bp) repeats within the SV40 enhancer. One centres around the core sequence identified by Weiher *et al.*¹³ and the second covers a region referred to here as the octamer motif. Proximal to the 21-bp repeats a third footprint was found which corresponds to the P-box¹⁴. We observed a similar pattern in the octamer region with extracts from Molt-4 and HeLa cells, and the entire pattern of protection with Molt-4 extract resembled that obtained¹⁵ with HeLa extracts.

The SV40 enhancer represents the prototype of a ubiquitously acting sequence element that stimulates RNA polymerase II transcription in nearly all cell types tested¹⁶. In contrast, the enhancer of the mouse immunoglobulin heavy-chain gene (IG) is highly tissue-specific^{17,18}. When a *Pst*I-*Eco*RI segment which contains most of the IG enhancer activity¹⁷ was taken for DNase I footprinting analysis three major areas of nuclease protection were detected (Fig. 1a, IG). These apparent protein-binding sites map to regions that have been found to be insensitive to DMS methylation *in vivo*¹⁹, indicating that the DNase protection observed *in vitro* represents sites of factor-DNA interaction *in vivo*.

The U snRNA genes of *Xenopus laevis* are developmentally regulated²⁰⁻²³. Two elements have been identified in the 5' flanking region of the U2 snRNA gene which are essential for efficient transcription²⁴. The distal sequence element (DSE) is located around position -260, and the proximal one (PSE) around position -50, with respect to the RNA start site. The footprint over the DSE is shown in Fig. 1a, U2.

Sequence comparison reveals similarities to the octamer motif of the SV40 enhancer in the IG enhancer between nucleotides 540 and 550 (ref. 19) and in the U2 DSE (Fig. 1b). Related sites have also been identified in the promoter regions of IG heavy- and κ light-chain genes^{3,4}, different H2b promoters²⁵ and U1 and U2 genes^{22,26-29}.

The IG and U2 enhancers contain 10 bp that are identical (5' ATGCAAATAG 3'), the first 7 bp also being conserved in SV40, and DNase protection and hypersensitive sites on both strands are found at comparable locations, relative to the common sequence motif, in all three cases (Fig. 1b). The possibility that the three DNA domains are recognized by the same component was examined in conditions where specific binding of the octamer binding factor could be detected. Figure 2a shows an experiment in which an end-labelled fragment containing the U2 DSE was incubated with nuclear extract and then run on a 5% native polyacrylamide gel. Three bands are seen, corresponding to the free DNA fragment and two species of retarded mobility (x and x'). To prove that the x and x' complexes resulted from specific binding a combined DNase I-fragment retardation experiment was carried out (Fig. 2b). This showed that the region of the retarded bands protected against DNase I digestion correspond to the DSE (compare Fig. 2b, lanes 3 and 4, with Fig. 1a, U2 and Fig. 4b, U2). The protection in the 3' (upper) region in this experiment is reduced compared to other, similar, experiments and to the protection observed under normal footprinting conditions (Figs 1, 4). Analogous results were obtained with restriction fragments of the SV40 and IG enhancer (not shown) and have also been reported for an (Ig) κ -immunoglobulin chain promoter fragment containing a sequence similar to the octamer motif⁵.

We next performed a competition assay to show that the same factor binds to the octamer motifs of the three enhancers. Double-stranded oligonucleotides containing the octamer motif regions of the three enhancers as well as a control oligonucleotide, were used in DNA fragment retardation assays with a fractionated HeLa cell nuclear extract (Fig. 3). All three enhancer oligonucleotides gave rise to complexes with similar or identical mobility (Fig. 3, U2, SV40, IG, lane 0). We then carried out binding assays in the presence of increasing amounts of unlabelled competitor DNA corresponding to either the U2 DSE (Fig. 3, lanes 1-4), the control oligonucleotide, which does not give rise to specific complexes with this fraction of the HeLa cell extract (Fig. 3, lanes 5-8), the SV40 enhancer octamer-motif region (Fig. 3, lanes 9-12) or the IG enhancer (Fig. 3, lanes 13-16). The three enhancer oligonucleotides compete with the labelled probe for complex formation, whereas the control oligonucleotide does not. This indicates that all three enhancer fragments bind to the same factor in the nuclear extract. The SV40 oligonucleotide competes less well than the IG or U2 fragments, suggesting that it has a lower affinity for the factor, perhaps because the SV40 enhancer octamer domain is less similar (Fig. 1b).

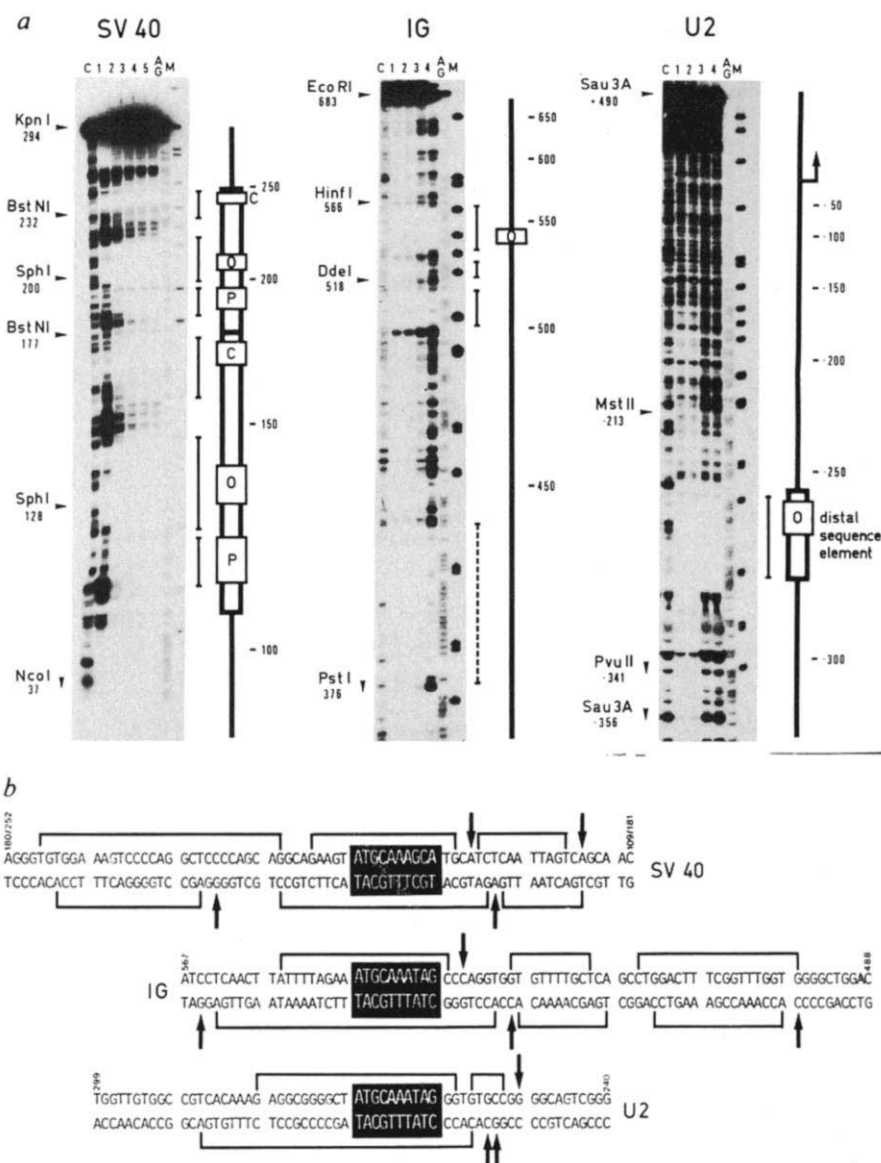
To purify octamer binding factor further, we used chromatographic fractionation (Fig. 4a). As measured by the fragment retardation assay the binding activity is mainly in fraction PC 350 (Fig. 4a), which was used for a footprinting experiment. The octamer motifs on the three enhancers showed the same protection pattern as with nuclear extracts whereas the other footprints seen with total extract (Fig. 1a) were highly reduced, showing that the factor binding to the octamer motif had been enriched relative to the other enhancer binding factors.

The binding activity found in the PC 350 fraction should be correlated with enhancer function. The stimulatory effect of the SV40 enhancer on transcription *in vitro* can be suppressed by increased template concentrations⁸, perhaps because of a distribution of promoter and enhancer factors, which have to act simultaneously, to different template molecules. Under these conditions the transcription efficiency should be increased by the addition of enhancer-binding factors only if the template

* Present addresses: Imperial Cancer Research Fund Laboratories, London WC2A 3PX, UK (T.D.); Zentrum für Molekulare Biologie, University of Heidelberg D-6900 Heidelberg, FRG (H.R.S.)

Fig. 1 *a*, DNase protection on SV40, IG and U2 enhancer and promoter fragments in unfractionated Molt-4 nuclear extract. Digestion products of control reactions without protein (C) and footprinting reactions (1–5) were run together with an A+G-specific sequencing lane (A+G) and an end-labelled *Hpa*II digest of pBR322 (M). Different DNase I dilutions were added for digestion (see below). Important restriction sites are indicated on the left of the autoradiographs. Footprints detected and descriptions of the end-labelled DNA fragments are shown diagrammatically on the right: P, P-box; O, octamer motif; C, enhancer core. SV40 (early coding strand): The 258 nucleotide (nt) *Nco*I–*Kpn*I fragment³² from position 38 to 294 was 5' labelled at the *Nco*I site with [γ -³²P]ATP. IG (non-coding strand): the fragment was 5'-labelled 46 bp away from the *Pst*I site at position 376 in a polylinker *Hind*III site and cut at the *Eco*RI site at position 683. Although the footprint between position 400 and 450 on the IG enhancer is not very obvious with this fragment, experiments with DNA fragments from this region labelled at different sites confirmed its presence (not shown). U2 (non-coding strand): the fragment was labelled at a polylinker *Eco*RI site 10 nt away from the *Sau*3A site at position –359 and cut at a polylinker *Pst*I site downstream of the gene. *b*, Comparison of the DNase I protection patterns around the octamer motif homologies in the SV40, the IG and the U2 enhancers. A schematic representation of footprinting experiments performed with different DNA fragments (coding and non-coding strands) is shown. Related sequence motifs (white letters) are positioned above each other. Brackets, protected areas; arrows, DNase hypersensitive sites. Numbers at the boundaries, position of presented sequences.

Methods. Nuclear extracts were prepared according to Dignam *et al.*³³ with modifications⁹. A+G sequencing reactions followed the standard protocol³⁴. Nuclear extract (4 μ l, 33 μ g protein) was preincubated with 25 ng of linearized pUC plasmid DNA on ice for 15 min. Reactions (10 μ l) contained 10 mM HEPES-KOH (pH 7.9), 20 mM KCl, 4 mM MgCl₂, 4 mM spermidine, 0.1 mM EDTA, 0.25 mM DTT and 10% glycerol. End-labelled DNA (10,000 c.p.m.) was added in 2 μ l H₂O and incubated for 10 min at 20 °C. DNase I (2 μ l, Worthington DPFF), freshly diluted in the incubation buffer to concentrations between 50 and 800 μ g ml⁻¹, was added and the reactions incubated for 90 seconds at 20 °C. The DNase concentrations were: Controls without extract (C lanes), 100 μ g ml⁻¹; SV40 lane 1, 800 μ g ml⁻¹; lane 2, 400 μ g ml⁻¹; lane 3, 200 μ g ml⁻¹; lane 4, 100 μ g ml⁻¹; IG and U2 lane 1, 100 μ g ml⁻¹; lane 2, 200 μ g ml⁻¹; lane 3, 400 μ g ml⁻¹; lane 4, 800 μ g ml⁻¹. In control reactions, extract was substituted by incubation buffer. DNase digestion was stopped by addition of 100 μ l phenol/chloroform/isoamylalcohol (25:24:1) and 90 μ l of 0.4 M LiCl containing 10 μ g transfer RNA. The aqueous phases were re-extracted with chloroform/isoamylalcohol (24:1), ethanol precipitated and loaded onto 6% acrylamide-urea sequencing gels.



contains a functional enhancer. Truncated β -globin templates either with (CAJ 6) or without (CAD, ref. 7) the SV40 enhancer were used for *in vitro* transcription (Fig. 4c). With high template concentrations we observed only a modest (twofold) enhancer activity (Fig. 4c, lanes 1 and 2). After addition of fraction PC 350 the amount of correctly initiated globin RNA was increased significantly only with the enhancer-containing template CAJ6 (Fig. 4c, lanes 3–6), resulting in an eightfold enhancer effect. Thus fraction PC 350, which contains the octamer-binding factor, does indeed stimulate transcription in an enhancer-dependent fashion. Although at this stage the factor is far from pure, it is likely that the DNA-binding activity and the factor stimulating enhancer-dependent *in vitro* transcription are the same, although additional enhancer binding factors are probably required to obtain the maximal *in vitro* effect.

The fact that the octamer motif was found to be part of the control regions of different transcription units such as the SV40,

IG and U2 genes and that it interacts, at least *in vitro*, with a common factor in these three cases indicates that this putative module cannot by itself explain the different regulatory characteristics of the three genes controlled by these enhancers. To explain how the different regulatory programmes are effected, additional steps in the control pathway must be assumed. In the case of the cell-specific IG expression there are indeed indications of tissue-specific factors binding to certain parts of the enhancer (ref. 30 and H.S. and D.B., unpublished data). Interactions with such proteins could then alter either the binding specificity or the transcriptional function of the octamer-binding protein. Alternatively, the octamer-binding factor might not work ubiquitously, being modified or replaced by a related, but cell- or stage-specific factor, necessary to allow proper functioning of regulated enhancers.

Singh and co-workers⁵ have demonstrated the presence in nuclear extracts from B cells and HeLa cells of a binding protein

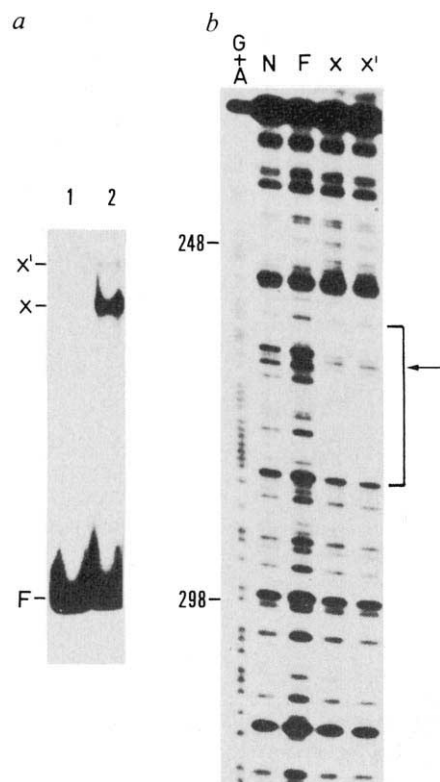


Fig. 2 Analysis of HeLa protein-U2 DSE complexes. *a*, A 5' end-labelled 150-bp *EcoRI*-*MstII* fragment containing the U2 DSE was incubated with HeLa nuclear extract, and the resulting complexes were analysed on a non-denaturing polyacrylamide gel. Lane 1, naked *EcoRI*-*MstII* fragment alone; F, mobility of the free DNA. Lane 2, after incubation with extract two retarded bands, x and x' are observed. Similar complexes are obtained with fragments from the IG and SV40 enhancer which contain the octamer motif. *b*, Analysis of regions protected against DNase I digestion in the x and x' complexes. The *EcoRI*-*MstII* fragment containing the U2 DSE was incubated with HeLa extract, digested with DNase I, then complexes were separated on a non-denaturing polyacrylamide gel. Bands corresponding to F, x and x' were excised, DNA was eluted and then run on a denaturing polyacrylamide-urea gel. Lane G+A, G+A sequence of the *EcoRI*-*MstII* fragment; lane N, naked DNA digested with 5 $\mu\text{g ml}^{-1}$ DNase I; lanes F, x and x', DNA eluted from the corresponding bands. Bracket, region of protection in the retarded fragment; arrow, position of increased DNase I sensitivity. In other experiments the protection extending upwards from this point was greater. Numbered residues correspond to those shown in Fig. 1b. **Methods.** *a*, End-labelled fragments (3,000 c.p.m.) containing the putative protein-binding site were incubated together with HeLa nuclear extract (0.5 μl , 4 μg protein) and 1.5 μg poly[d(I-C)] in 15 μl reaction mix. Buffer was 10 mM HEPES-KOH (pH 7.9), 100 mM KCl, 4 mM MgCl_2 , 4 mM spermidine, 0.1 mM EDTA, 0.25 mM dithiothreitol (DTT), 100 $\mu\text{g ml}^{-1}$ bovine serum albumin and 10% glycerol. Samples were incubated at 25°C for 10 min to allow binding. Loading buffer, (40% glycerol, 50 mM EDTA, 5 $\times$ TBE, 0.25% bromophenol blue) was added (3.8 μl), and samples loaded onto a 5% non-denaturing acrylamide gel (acrylamide/bisacrylamide 30:1). Electrophoresis was carried out in 1 $\times$ TBE at 10 V cm^{-1} . Conditions were similar to those described previously^{3,35-37}. *b*, Conditions similar to *a*, but probe (300,000 c.p.m.), 8 μl extract, 6 μg poly[d(I-C)] and 20 mM KCl in a final volume of 21 μl was used. After binding DNase I was added to 50 $\mu\text{g ml}^{-1}$ and digestion was allowed to proceed for 90 s before loading buffer was added. DNA was eluted from the gel³⁴, resuspended and loaded onto a 6% polyacrylamide-urea gel.

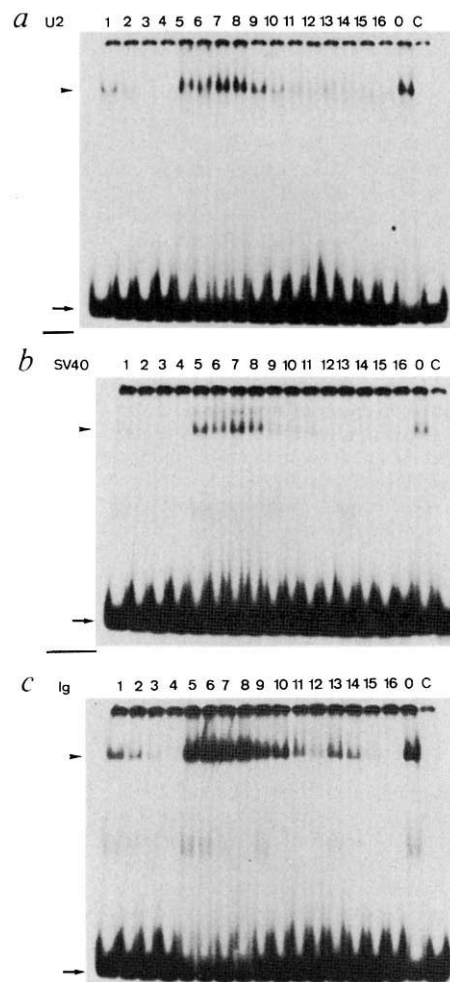
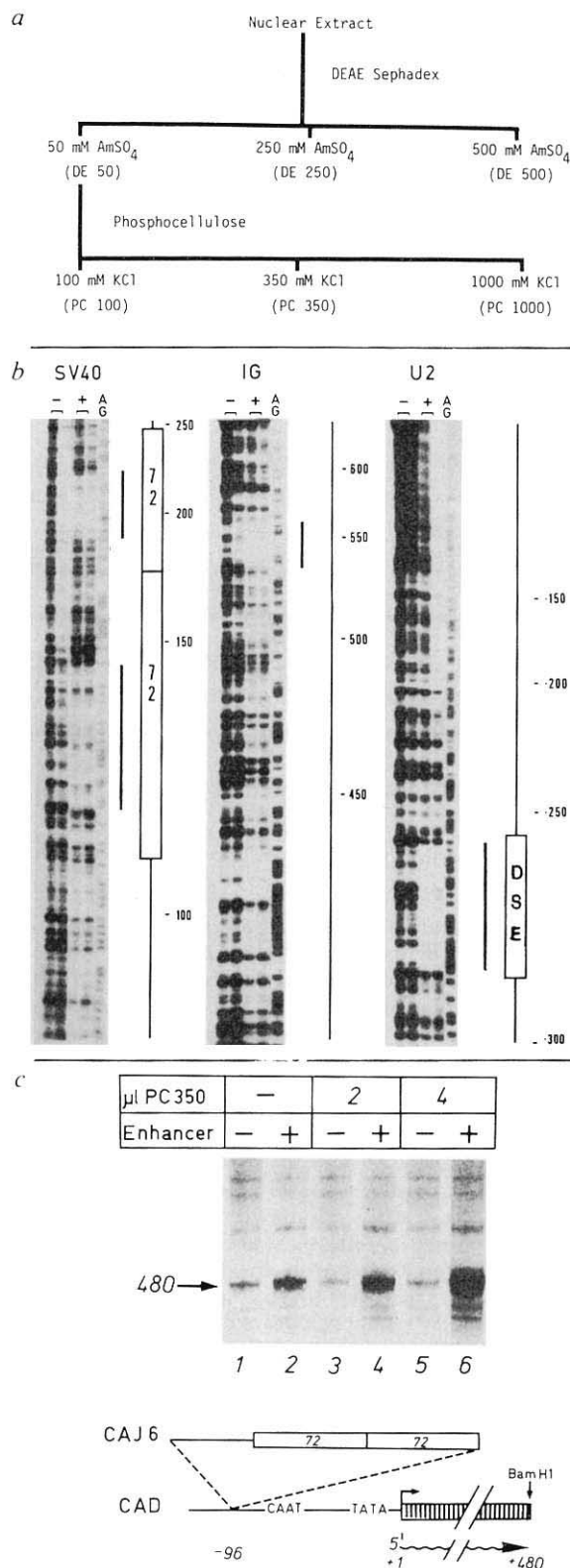


Fig. 3 The three enhancer regions contain binding sites for the same factor. Four double-stranded 40-mer oligonucleotides were synthesized, one corresponding to the U2 DSE region (from position 288-249 (Fig. 1b), one to part of the SV40 enhancer (152/224-113/185), one to part of the IG enhancer (567-528) and another control oligonucleotide which did not contain an octamer motif homology. These were then used to compete with one another for binding of factors present in a HeLa cell nuclear extract. *a*, U2 DSE oligonucleotide (3,000 c.p.m., 2 pg) was used in the assay. The major retarded complex is indicated with an arrowhead. Lane C, naked DNA control; lane O, no competitor; lanes 1-4, 50, 100, 200 and 500-fold molar excess of unlabelled U2 DSE oligonucleotide were added as competitor; lanes 5-8: 50, 100, 200 and 500-fold molar excess of unlabelled control oligonucleotide were added as competitor; lanes 9-12: 50, 100, 200 and 500-fold molar excess of the SV40 oligonucleotide were added as competitor; lanes 13-16: 50, 100, 200 and 500-fold molar excess of the IG oligonucleotide were added as competitor. *b*, Analogous to *a*, but with the 5' end-labelled SV40 oligonucleotide as probe. All competitor concentrations halved with respect to *a*. *c*, Analogous to *a*, but with the 5' end-labelled IG oligonucleotide as probe. Competitor concentrations as in *b*.

Methods. DNA fragment retardation assays were carried out as described in Fig. 2, except that a HeLa extract was used which had been fractionated on DEAE Sephadex-A25 (see Fig. 4), then bound to heparin-Sepharose in 100 mM KCl and eluted in a step of 500 mM KCl to enrich for the octamer motif binding factor. Competitor and labelled oligonucleotides were added simultaneously following preincubation of the extract with poly[d(I-C)] as in Fig. 2.

Fig. 4 Partial purification of the octamer binding factor. *a*, Fractionation scheme; *b*, DNase I footprinting assay with fraction PC 350 on SV40, IG and U2 fragments. DNA and footprinting conditions were the same as described in Fig. 1 legend except that $30 \mu\text{g ml}^{-1}$ of poly[d(I-C)] were used as unspecific competitor. Concentrations of DNase were (from left to right in each panel): 12 and $50 \mu\text{g ml}^{-1}$ for the control lanes (–) and 50 and $100 \mu\text{g ml}^{-1}$ for the footprinting lanes (+). For each reaction $10 \mu\text{l}$ PC 350 were used. Major areas protected against nuclease are indicated by solid lines. Next to each experiment an A + G-specific sequencing lane was run. DSE, U2 distal sequence element. *c*, Stimulation of β -globin transcription *in vitro* by the PC 350 fraction. In plasmid CAJ6 the SV40 enhancer from position 100 to 296 is inserted 96 bp upstream of the β -globin cap-site in its wild-type orientation⁷. The templates were cleaved at a *Bam*HI site in the coding region as shown in the lower part of the figure. The expected length of the run-off transcript is 480 nucleotides. Transcriptions were done in the absence or presence of PC 350 as indicated.

Methods. *a*, HeLa nuclear extract prepared by the standard protocol was dialysed against DE-buffer (20 mM HEPES-KOH (pH 7.9), 50 mM $(\text{NH}_4)_2\text{SO}_4$, 2 mM MgCl_2 , 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol) and loaded onto a DEAE-Sephadex A-25 column. The column was washed with loading buffer and eluted with DE-250 and DE-500 buffer (DE-buffer with 250 and 500 mM $(\text{NH}_4)_2\text{SO}_4$, respectively). Proteins in the flow-through were precipitated with $(\text{NH}_4)_2\text{SO}_4$, 0.38 g ml^{-1} . The protein pellet was taken up in one tenth the original volume of PC-loading buffer (20 mM HEPES-KOH, (pH 7.9), 100 mM KCl, 0.5 mM DTT, 20% glycerol) and dialysed against this buffer. This material was passed onto a phosphocellulose column (Whatman P11), washed and eluted with PC 350 and PC 1000-buffers (PC-loading buffer with 350 mM and 1 M KCl, respectively). Flow-through and step-fractions were dialysed against the standard extract buffer D* (20 mM HEPES-KOH (pH 7.9), 20 mM KCl, 2 mM MgCl_2 , 0.2 mM EDTA, 0.5 mM DTT, 20% glycerol. Protein concentrations: PC 100, 0.72 mg ml^{-1} ; PC 350, 0.92 mg ml^{-1} ; PC 1000, 0.25 mg ml^{-1} . *b*, Transcription reactions contained (20 μl) 4 μl HeLa nuclear extract, 0–4 μl PC 350, 1 μg truncated DNA template, 185 kBq [α - ^{32}P]UTP (15 TBq mmol^{-1}); 5 mM creatine phosphate; 0.5 mM ATP, 0.5 mM CTP, 0.5 mM GTP, 10 μM UTP, 20 mM KCl, 4 mM MgCl_2 , 4 mM spermidine, 10 mM HEPES-KOH (pH 7.9), 0.1 mM EDTA, 0.25 mM dithiothreitol, 10% glycerol. The samples were incubated at 30°C for 60 min and then extracted with phenol/chloroform/isoamylalcohol (25:24:1). The RNAs were precipitated with 2.5 M ammonium acetate and 2.5 vols ethanol, separated on a 6% acrylamide-urea sequencing gel, and detected by autoradiography. Only the part of the autoradiograph containing the specific run-off RNA band (arrow) is shown.



(IgNF-A) specific for the octamer motif in the promoter of an immunoglobulin κ light-chain gene whose binding could be competed by DNA fragments from the IG heavy-chain enhancer. It is therefore likely that IgNF-A and the octamer-binding factor are identical, although the IgNF-A footprint seems to be smaller than the protected regions we detect on several octamer motif regions.

Preliminary results from comparative footprinting experi-

ments (in collaboration with G. J. M. Pruijn, W. van Driel and P. van der Vliet) indicate that nuclear factor III, a protein isolated from HeLa cell nuclei which is required for adenovirus DNA replication in an *in vitro* system³¹, recognizes the same DNA sequence as the octamer-binding factor described here.

We thank Werner Fleischer and Nina Dathan for technical assistance, and Brian Sproat and Phillipe Neuner for synthesizing the oligonucleotides. We also thank David Friendewey,

Timothy Humphrey and Angela Krämer for comments on the manuscript and Alain Sergeant for constructive criticism. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie to W.K. **Note added in proof:** Recent experiments³⁸ have shown that restriction fragments from the promoters of human U1 snRNA, histone H2B and an immunoglobulin κ -chain gene, all of which contain an octamer motif, compete for the binding of a single factor in HeLa cell nuclear extracts. The octamer motif is required for maximal *in vitro* transcription of the histone H2B gene³⁹.

Received 21 July; accepted 11 November 1986.

- Serfling, E., Jasin, M. & Schaffner, W. *Trends Genet.* **1**, 224-230 (1985).
- Voss, D., Schlokot, U. & Gruss, P. *Trends Biochem.* **11**, 287-289 (1986).
- Falkner, F. G. & Zachau, H. G. *Nature* **310**, 71-74 (1984).
- Parslow, T. G., Blair, D. M., Murphy, W. J. & Granner, D. K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2650-2654 (1984).
- Singh, H., Sen, R., Baltimore, D. & Sharp, P. A. *Nature* **319**, 154-158 (1985).
- Sen, R. & Baltimore, D. *Cell* **46**, 705-716 (1986).
- Sergeant, A., Bohmann, D., Weiher, H., Zentgraf, H. & Keller, W. *J. molec. Biol.* **180**, 577-600 (1984).
- Sassone-Corsi, P., Dougherty, J. P., Wasylyk, B. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **81**, 308-312 (1984).
- Wildeman, A., Sassone-Corsi, P., Grundström, T., Zenke, M. & Chambon, P. *EMBO J.* **3**, 3129-3133 (1984).

- Schöler, H. & Gruss, P. *EMBO J.* **4**, 3005-3013 (1985).
- Galas, D. J. & Schmitz, A. *Nucleic Acids Res.* **5**, 3157-3170 (1978).
- Minowada, J., Ohnura, T. & Moore, G. E. *J. natn. Cancer Inst.* **49**, 891-895 (1972).
- Weiher, H., König, M. & Gruss, P. *Science* **219**, 626-631 (1983).
- Zenke, M. *et al.* *EMBO J.* **5**, 387-397 (1986).
- Wildeman, A. G. *et al.* *Molec. cell. Biol.* **6**, 2098-2105 (1986).
- Mosthaf, L., Pawlita, M. & Gruss, P. *Nature* **315**, 597-600 (1985).
- Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729-740 (1983).
- Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 718-728 (1983).
- Ephrussi, A., Church, G. M., Tonegawa, S. & Gilbert, W. *Science* **227**, 134-140 (1985).
- Zeller, R., Nyffenegger, T. & De Robertis, E. M. *Cell* **32**, 425-434 (1983).
- Forbes, D. J., Kirschner, M. W., Caput, D., Dahlberg, J. E. & Lund, E. *Cell* **38**, 681-689 (1984).
- Fritz, A., Parisot, R., Newmeyer, D. & De Robertis, E. M. *J. molec. Biol.* **178**, 273-285 (1984).
- Mattaj, J. W. *et al.* *Oxf. Surv. Eucaryotic Genes* **2**, 121-140 (1985).
- Mattaj, J. W., Lienhard, S., Jiricny, J. & De Robertis, E. M. *Nature* **316**, 163-167 (1985).
- Harvey, R. P., Robins, A. J. & Wells, J. R. E. *Nucleic Acids Res.* **10**, 7851-7863 (1982).
- Ciliberto, G., Buckland, R., Cortese, R. & Philipson, L. *EMBO J.* **4**, 1537-1543 (1985).
- Szukeski, J. M. *et al.* *J. biol. Chem.* **259**, 8345-8352 (1984).
- Krol, A., Lund, E. & Dahlberg, J. E. *EMBO J.* **4**, 1529-1535 (1985).
- Ares, M. Jr., Mangin, M. & Weiner, A. M. *Molec. cell. Biol.* **5**, 1560-1570 (1985).
- Wasylyk, C. & Wasylyk, B. *EMBO J.* **5**, 553-560 (1986).
- Prujin, G. J. M., van Driel, W. & van der Vliet, P. *Nature* **322**, 656-659 (1986).
- Toozé, J. (ed.) *DNA Tumor Viruses* (Cold Spring Harbor Laboratory, New York, 1980).
- Dignam, J. D., Lebowitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475-1489 (1983).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Fried, M. & Crothers, D. M. *Nucleic Acids Res.* **9**, 6505-6525 (1981).
- Piette, J., Kryszke, M. H. & Yaniv, M. *EMBO J.* **4**, 2675-2685 (1985).
- Schneider, R., Gauder, I., Müller, U., Mertz, R. & Winnacker, E. L. *Nucleic Acids Res.* **14**, 1303-1317 (1986).
- Sive, H. L. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6382-6386 (1986).
- Sive, H. L., Heintz, N. & Roeder, R. G. *Molec. cell. Biol.* **6**, 3329-3340 (1986).

Expression of a polypeptide containing a dipeptide repeat is confined to the insect stage of *Trypanosoma brucei*

Isabel Roditi*, Mark Carrington* & Mervyn Turner*

MRC Biochemical Parasitology Unit, Moltano Institute, Downing Street, Cambridge CB2 3EE, UK

The protozoan parasite *Trypanosoma brucei* is transmitted between mammalian hosts by the tsetse fly (*Glossina* spp.). Trypanosomes ingested by the fly undergo a number of changes in the insect midgut during differentiation to procyclic forms. These include the loss of the variant specific glycoprotein (VSG) coat¹ and the appearance of a common set of procyclic surface antigens². In order to investigate genes other than VSG genes which are expressed only at certain stages of the life cycle, the first cDNA specific to procyclic culture form trypanosomes (equivalent to the stage found in the insect midgut) has been characterized. The encoded polypeptide shows several characteristics of membrane proteins, but its most striking feature is the presence of a repetitive amino-acid sequence in which there are 22 tandem repeats of the dipeptide -Glu-Pro-. Related genes are also found in other trypanosome species and in leishmania. This gene shows many similarities to a number of surface antigen genes described in malaria³ and, more recently, *Trypanosoma cruzi*⁴. This is the first example of a repetitive sequence in a parasite protein which is present only in the insect vector, and which therefore cannot be implicated in the mammalian host immune response⁴⁻⁶.

A complementary DNA library was constructed from messenger RNA purified from procyclic culture forms of *T.b. brucei* 427.01 (ref. 7). The library was screened with single-stranded cDNAs synthesized from procyclic and bloodstream mRNAs. Of the plaques that hybridized differentially to these probes, the majority showed only quantitative differences; relatively few displayed absolute stage specificity. This is consistent with the limited number of stage-specific polypeptides among *in vitro* translation products (data not shown). This approach contrasts with methods previously used to characterize stage-regulated gene expression in trypanosomes, which have employed heterologous probes such as *Drosophila* heat shock

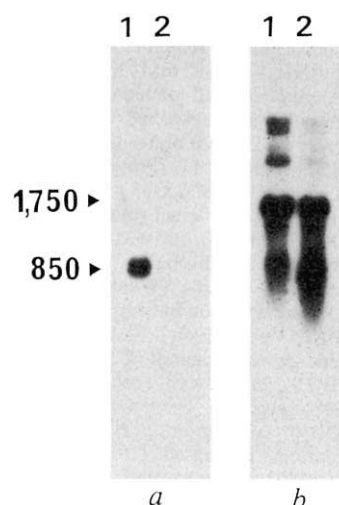


Fig. 1 Northern blot analyses of *T.b. brucei* procyclic culture form and bloodstream form RNAs from the MITaR serodeme⁷. Total RNA was glyoxalated²⁵ then separated on a 1.3% agarose gel, followed by transfer to nitrocellulose. Lane 1, 5 μ g procyclic culture form RNA (427.01); lane 2, 5 μ g bloodstream form RNA (MITaR 1.6). The blot was probed first with (a) nick-translated ³²P-labelled pPRO2001 (ref. 26), then reprobed with (b) pPRO301. Size markers in base pairs.

Methods. RNA was isolated from bloodstream and procyclic trypanosomes using the phenol-SDS method²⁷. Poly(A) RNA was selected by oligo(dT) chromatography²⁸. Double stranded cDNA was synthesized by the strand replacement method²⁹, prepared for vector ligation using *Eco*RI linkers³⁰, and inserted into the *Eco*RI site of λ NM1149 and a portion of the unamplified library plated on *E. coli* NM514³⁰. The screening was performed by plaque hybridization³¹ using ³²P-labelled single-stranded cDNA synthesized by AMV reverse transcriptase using poly(A) RNA from either bloodstream form or procyclic trypanosomes as template and random hexadeoxynucleotides as primers³². Plaques that hybridized with procyclic but not bloodstream cDNA were purified. DNA was isolated from one of these³³ and the cDNA insert subcloned into the *Eco*RI site of pUC8³⁴ to produce pPRO2001. A second clone, pPRO301, which did not show stage specificity, was isolated from the same library for use as a positive control.

genes⁸. The first stage-specific clone to be characterized (pPRO2001) hybridized to an mRNA of ~850 bases found in procyclic, but not bloodstream form, trypanosomes (Fig. 1a). Even on very long exposure of the autoradiograph there was

* Present addresses: Institut für Genetik, Kernforschungszentrum, Postfach 3640, 7500 Karlsruhe 1, FRG (I.R. & M.C.) and Merck, Sharp & Dohme Research Laboratories, PO Box 2000, Rahway, New Jersey 07065-0900, USA (M.T.)

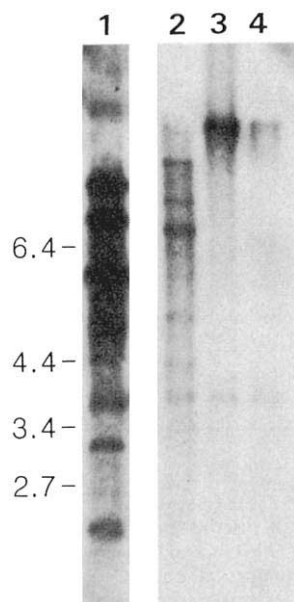


Fig. 4 Southern blot analysis of genes related to pPRO2001 in other Kinetoplastida. Genomic DNA (1 µg) was digested with *Pst*I, size fractionated on a 0.8% agarose gel and transferred to nitrocellulose³⁷. After hybridization with nick-translated pPRO2001²⁶ the blot was washed with a solution of 0.1×SSC, 0.1% SDS at 60 °C for 1 h before autoradiography for 4 d. Lane 1, *T. b. brucei* (427.01); lane 2, *T. congolense* (ILNat 2.1); lane 3, *T. vivax* (1223); lane 4, *L. m. amazonensis* (PH8). Size markers, kb.

form variants that had been cloned by syringe passage (Fig. 3). Interestingly, in a procyclic culture derived from one of these bloodstream clones, polymorphism in one fragment was observed with several restriction enzymes. The serodeme MITaR 1, isolated in Uganda⁷, shows a polymorphism which distinguishes it from the ILTaR 1 serodeme (Fig. 3), although no rearrangement was found between a pair of bloodstream and procyclic forms within this serodeme. Certain regions of the trypanosome genome are inherently less stable than others. Both VSG genes and their expression-site associated genes²¹ are prone to rearrangement, whereas tubulin genes and various glycolytic enzyme genes are much more stable²². The genes encoding the procyclic form specific polypeptide seem to belong to the former class.

Sequences showing a high degree of homology to these genes are also present in other members of the order Kinetoplastida. They can be detected in *Trypanosoma congolense*, *T. vivax* and *Leishmania mexicana amazonensis* (Fig. 4), although the reduced intensity of hybridization suggests that the genes are not exact copies of those in *T. brucei*. Most characterized polypeptides containing tandem repeats occur in protozoa, although there are exceptions such as the glue proteins of *Drosophila*^{23,24}. Those identified in *Plasmodium* and *T. cruzi* are antigens which evoke an immune response during the course of infection^{3,4}. The only antigens of *T. brucei* which have been well characterized are VSGs and these do not contain any amino-acid repeats. It is significant that these are the surface antigens expressed by trypanosomes which have to contend with the mammalian immune system, whereas the procyclic form specific polypeptide containing amino-acid repeats is expressed by a life cycle stage which is confined to the insect. This finding argues against tandem amino-acid repeats having originally evolved in parasitic protozoa to provide protection exclusively in the mammalian host.

We thank Judith Creighton for bloodstream form trypanosomes, Alistair Ross and Jerry Wells for procyclic forms, Phelix Majiwa for gifts of *T. congolense* and *T. vivax* DNA, Lorna

Gibson for *L. m. amazonensis* DNA, Dick Williams, Dirk Döbelare and Amanda Gill for criticizing the manuscript, and Josie McConnell and Angelika Ehrfeld for advice and encouragement.

Received 24 October; accepted 24 November 1986.

- Barry, J. D. & Vickerman, K. *Expl. Parasit.* **48**, 313–324 (1979).
- Seed, J. R. *Parasitology* **54**, 593–596 (1964).
- Kemp, D. J. *et al. Parasitology* **91**, S83–S108 (1986).
- Peterson, D. S., Wrightsman, R. A. & Manning, J. E. *Nature* **322**, 566–568 (1986).
- Ravetch, J. V., Kochan, J. & Perkins, M. *Science* **227**, 1593–1597 (1985).
- Berzins, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 1065–1069 (1986).
- Cross, G. A. M. & Manning, J. C. *Parasitology* **67**, 315–331 (1973).
- Van der Ploeg, L. H. T., Gianni, S. H. & Cantor, C. R. *Science* **228**, 1443–1445 (1985).
- De Lange, T., Michels, P. A. M., Veerman, H. J. G., Cornelissen, A. W. C. A. & Borst, P. *Nucleic Acids Res.* **12**, 3777–3790 (1984).
- Parsons, M., Nelson, R. G., Watkins, K. P. & Agabian, N. *Cell* **38**, 309–316 (1984).
- Clayton, C. E. *EMBO J.* **4**, 2997–3003 (1985).
- Boothroyd, J. C., Paynter, C. A., Cross, G. A. M., Bernards, A. & Borst, P. *Nucleic Acids Res.* **9**, 4735–4743 (1981).
- Kozak, M. *Cell* **44**, 283–292 (1986).
- Blobel, G. & Dobberstein, B. *J. cell Biol.* **67**, 835–851 (1975).
- Porter, A. G. *et al. Nature* **282**, 471–477 (1979).
- Ward, C. W. & Dopheide, T. A. *Br. med. Bull.* **35**, 51–56 (1979).
- Williams, R. O., Young, J. R. & Majiwa, P. A. O. *Nature* **299**, 417–421 (1982).
- De Lange, T. & Borst, P. *Nature* **299**, 451–453 (1982).
- Van der Ploeg, L. H. T., Schwartz, D. C., Cantor, C. R. & Borst, P. *Cell* **37**, 77–84 (1984).
- Miller, E. N. & Turner, M. J. *Parasitology* **82**, 63–80 (1981).
- Cully, D. F., Ip, H. S. & Cross, G. A. M. *Cell* **42**, 173–182 (1985).
- Gibson, W. C., Osinga, K. A., Michels, P. A. M. & Borst, P. *Molec. biochem. Parasit.* **16**, 231–242 (1985).
- Muskavitch, M. A. T. & Hogness, D. S. *Cell* **29**, 1041–1051 (1982).
- Garfinkel, M. D., Pruitt, R. E. & Meyerowitz, E. M. *J. molec. Biol.* **168**, 765–789 (1983).
- McMaster, G. K. & Carmichael, G. G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4835–4838 (1977).
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
- Carrington, D. M., Auffret, A. D. & Hanke, D. E. *Nature* **313**, 64–67 (1985).
- Caton, A. J. & Robertson, J. S. *Nucleic Acids Res.* **7**, 1445–1456 (1979).
- Watson, C. J. & Jackson, J. F. in *DNA Cloning: A Practical Approach* Vol. 1 (ed. Glover, D.) 79–88 (IRL, Oxford, 1985).
- Murray, N. in *Lambda II* (eds Hendrix, R. W. *et al.*) 395–432 (Cold Spring Harbor Laboratory, New York, 1983).
- Grunstein, M. & Hogness, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6–13 (1983).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Vieira, J. & Messing, J. *Gene* **19**, 259–268 (1982).
- Messing, J. *Meth. Enzym.* **101**, 20–77 (1983).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Young, J. R., Donelson, J. E., Majiwa, P. A. O., Shapiro, S. Z. & Williams, R. O. *Nucleic Acids Res.* **10**, 803–819 (1982).

Transgenic rye plants obtained by injecting DNA into young floral tillers

A. de la Peña*†, H. Lörz & J. Schell

Max Planck Institut für Züchtungsforschung, D-5000 Köln 30, FRG

The most successful method currently used to produce transgenic plants is based on the ability of *Agrobacterium tumefaciens* to infect a number of higher plant species and transfer a defined DNA fragment (T-DNA) to the genome of the infected cells¹. This procedure cannot be used with the agriculturally important cereals, although preliminary data have suggested² infection of maize seedlings by *A. tumefaciens* strains. Purified exogenous DNA can be taken up, integrated and expressed in cells of a variety of plant species including some cereals following direct gene transfer into isolated protoplasts^{3–7}. However, although it is possible to grow isolated cereal protoplast into unorganized tissue (calli)^{5,7,8}, only rice protoplasts have so far been shown to regenerate mature plants^{9–11}. Here we report an alternative approach to transformation of cereal plants which does not involve tissue culture techniques.

Previous work^{12,13} on the development of the male germ line of rye (*Secale cereale* L.) has shown that at 14 days before the first meiotic metaphase the archesporial cells are highly sensitive

* Permanent address: Departamento de Genética, Facultad de Biología, Universidad Complutense, Madrid, Spain.

† To whom reprint request should be sent.

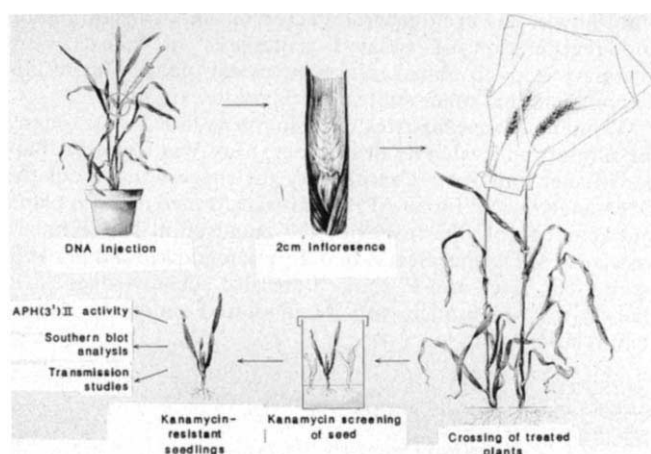


Fig. 1 Scheme of the transformation and screening strategy. Aliquots of an aqueous solution ($100 \mu\text{g ml}^{-1}$) of pLGVneo1103 were injected above each tiller node until several drops of the solution came out from the top of the tiller (usually a total injected volume of $300 \mu\text{l}$). The plants were injected about 14 days before meiosis using a tuberculin syringe (25G 5/8). This stage was recognized by the morphology of the developing tiller. When young tillers have 5 leaves and the flag leaf reaches one third to one half of its final size, they contain, at approximately the level of the third leaf, a young inflorescence of about 2 cm which is suitable for injection. After injection, the floral tillers were allowed to grow to maturity. Seeds from these injected tillers were obtained by cross-pollination with other injected tillers. This cross-pollination was necessary because JNK rye is strongly self-incompatible.

to caffeine and colchicine injected into the developing floral tillers. We considered that at this stage the archesporial cells might also be permeable to other molecules, such as DNA. Therefore we have injected DNA carrying a dominant selectable marker gene into rye plants, under precisely the same conditions as those leading to meiotic abnormalities after caffeine or colchicine injections.

The plasmid pLGVneo1103 (Fig. 3b) carrying the aminoglycoside phosphotransferase II gene (APH(3')II) under the control of the nopaline synthase (NOS) promoter (L. Herrera-Estrella, personal communication; ref. 4) was injected into developing floral tillers of the diploid rye cultivar JNK. The plants were injected 14 days before meiosis (Fig. 1). JNK rye expresses a high level of self-incompatibility, therefore, seeds were obtained by pair-wise crossing of injected tillers from different plants. The seeds set on the injected tillers were screened for kanamycin resistance. Glass containers (3l) were filled under aseptic conditions with about 400 ml of Knop nutritive solution^{14,15}, supplemented with $10 \mu\text{g ml}^{-1}$ of kanamycin sulphate (Sigma). A plastic net was positioned in these containers such that seeds deposited on it would lie just above, and in contact with, the nutritive solution. A maximum of 10 surface sterilized seeds were tested in each container and grown in a culture room at 26°C at approximately 2,000 lx for 16 h per day. Control seeds, from untreated rye plants, were germinated under the same conditions with and without kanamycin. In the presence of kanamycin, control seedlings bleached and were almost totally white after 10 days. From 3,023 seeds derived by cross-pollinating 98 plants injected with pLGVneo1103, which were tested with this procedure, seven seedlings remained green after 10 days growth on kanamycin-containing medium. These apparently kanamycin resistant seedlings were assayed for the presence of APH(3')II enzymatic activity^{16,17}. Two of these seven 'kanamycin resistant' plants, each resulting from an independent injection experiment, showed APH(3')II activity (Fig. 2, lanes f, h). No APH(3')II activity was detected in about one hundred control rye seedlings or in 30 bleached kanamycin-sensitive seedlings derived from injected tillers.

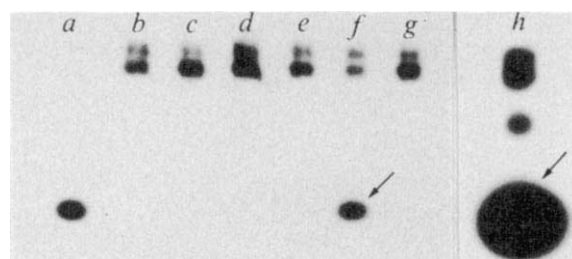


Fig. 2 Detection of APH(3')II activity in plant extracts. Leaf material ($\sim 100 \text{ mg}$) was crushed in an Eppendorf tube with $70 \mu\text{l}$ of extraction buffer¹⁷. Protein extracts ($50 \mu\text{l}$) were electrophoresed onto a 10% polyacrylamide gel, which was incubated with kanamycin and [$\gamma\text{-}^{32}\text{P}$]ATP and blotted onto Whatman phosphocellulose paper p81¹⁷. a, APH(3')II-transformed tobacco (positive control); b, Untreated JNK rye (negative control); c–e, untransformed kanamycin-sensitive seedlings from injected plants; f, kanamycin-resistant seedlings; g, apparently kanamycin resistant seedling showing no APH(3')II activity; h, APH(3')II activity detected in a second kanamycin-resistant seedling resulting from an independent experiment. Arrow, ^{32}P -labelled band of kanamycin that comigrates with one resulting from APH(3')II activity in a transformed tobacco plant provided by Dr P. Czernilofsky. The bands present in all the rye extracts have been observed in variable form for several plant species and are thought to represent protein kinases^{5,7,17}.

To confirm that the observed APH(3')II activity was due to the presence of the introduced pLGVneo1103 DNA in the plant genome, DNA/DNA hybridizations¹⁸ in which DNA isolated from the APH(3')II-positive plants, as well as from one APH(3')II-negative but apparently kanamycin resistant plant, were performed using a cloned 1.6 kilobase (kb) *Pst*I fragment of pLGVneo1103 internal to the APH(3')II gene as a specific probe (Fig. 3b). No hybridization was found in DNA isolated from control rye plants (lane d) while in the two APH(3')II-positive plants a 3 kb hybridizing band, corresponding to the *Eco*RI–*Hind*III fragment of pLGVneo1103, was detected (lanes f, h). Two additional bands, of 2.6 kb and 1.8 kb, were found in the DNA of one of the transformed plants (lane f) probably representing rearranged copies of the introduced gene, a common feature in gene transformation by direct DNA uptake^{4,7,19}. The DNA from one of the five 'kanamycin resistant' plants which did not show any APH(3')II activity did not contain any fragment hybridizing to the APH(3')II probe (lane i). This indicates that plants can remain green throughout the screening procedure without having incorporated the selectable marker although it reinforces the correlation between APH(3')II activity and actual DNA incorporation. The 'leaky' character of the kanamycin selection observed in JNK rye might be the consequence of the genotypic heterogeneity of this cultivar resulting from its cross-pollinating mating behaviour. Genomic DNAs digested separately with *Eco*RI or with *Hind*III, each cutting at a unique site in pLGVneo1103, showed several hybridizing bands (data not shown). This result is as expected if the introduced DNA is integrated in the genomic DNA at more than one site or as more than one copy, and the hybridizing bands larger than the 7.15 kb of pLGVneo1103 can also be explained in this way.

These results show that it is possible to introduce new genetic information into germ cells of cereals and to recover seeds capable of growing into normal plants in which the foreign gene is present and expressed. In a repeat experiment, 37 JNK plants were similarly injected with DNA of another plasmid carrying the same APH(3')II gene. Out of about 1,000 derived seeds, three were 'kanamycin resistant' and one of these showed APH(3')II activity. The occurrence of this third APH(3')II positive plant indicates the reliability of the transformation procedure. The method is based on delivery of DNA, carrying a

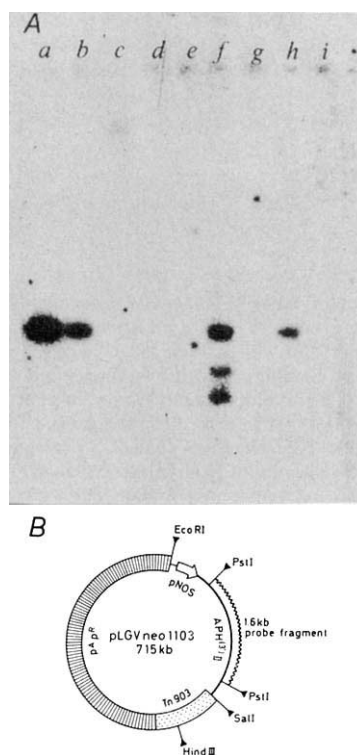


Fig. 3 DNA analysis of the APH(3')II positive plants. **A**, DNA was isolated from leaf material²³ and double digested with *Eco*RI and *Hind*III. Approx. 5 µg of digested DNA was analysed by electrophoresis in 0.8% agarose gel and blotted¹⁸. The probe was the 1.6 kb *Pst*I internal fragment of the APH(3')II gene. Lanes **a** and **b**, pLGVneo1103 DNA digested with *Eco*RI and *Hind*III (lane **a** contains an amount of pLGVneo1103 DNA approximately equivalent to 10 copies of plasmid DNA per haploid rye genome and lane **b** contains two copies equivalent); lane **c**, *Hind*III-digested λDNA, used for size marking of DNA fragments. Lane **d**, control JNK rye DNA digested with *Eco*RI and *Hind*III; lane **e**, DNA from APH(3')II positive plant 1; lanes **f**, **g**, blank. The arrow shows the expected 3 kb band comigrating with the diagnostic 3 kb *Eco*-*Hind* pLGVneo1103 fragment. Lane **h**, DNA from APH(3')II-positive plant 2. The arrow shows the expected 3 kb band. Lane **i**, DNA from apparently kanamycin resistant but APH(3')II-negative plant. **B**, Diagram of pLGVneo1103 showing the *Eco*RI and *Hind*III sites and the 1.6 kb *Pst*I fragment used as a probe.

selectable marker gene into young floral tillers. We postulate that the DNA is subsequently transported by the plant vascular system to the germ cells which can take it up only if they are in a 'competent' stage. In our rye cultivar this competent stage occurs two weeks before meiosis and corresponds cytologically to the interphase before the penultimate premeiotic mitosis in male tissue¹³, when male germ cells are not surrounded by a callose wall²¹. (At this time female cells are in an equivalent developmental stage²⁰). These cells undergo several changes in response to commitment to meiosis²², perhaps the most important in relation to DNA uptake being the establishment of cytoplasmic channels connecting the archesporial cells. It is conceivable that DNA molecules could pass through these structures²². At present we do not have any evidence to indicate whether DNA uptake occurred in the male or the female germ line or in both. The membrane characteristics of the male cells at the time of injection make them very likely candidates. Further experiments will test this hypothesis.

We are confident that this simple transformation procedure can be extended to other cereals, which have a premeiotic development equivalent to that of rye and for which, no other transformation system leading to plants is currently available.

Until now it has been generally accepted⁸ that transformation and regeneration of isolated protoplasts of cereals were necessary steps to obtain transgenic cereal plants. Our results demonstrate that other strategies are worth exploring.

We thank Javier Paz-Ares, Angelo Spina and Janusz Zimny for support and valuable discussions; Guy Van Canneyt, Barbara Baker and Peter Czernilofsky for suggestions about the DNA analysis and for an APH(3')II-transformed tobacco plant, and Peter Murphy for reviewing this manuscript. The technical assistance of Dagmar Becker and the photographic and drawing work of D. Bock and F. Furkert are also acknowledged. A.P. thanks the Alexander von Humboldt Foundation for a fellowship.

Received 3 September; accepted 28 October 1986.

1. Nester, E. W., Gordon, M. P., Amasino, R. M. & Yanofsky, M. F. *A. Rev. pl. Physiol.* **35**, G387-413 (1984).
2. Graves, A. C. F. & Goldman, S. L. *Pl. molec. Biol.* **7**, 43-50 (1986).
3. Paszkowski, J. *et al. EMBO J.* **3**, 2717-2722 (1984).
4. Hain, R. *et al. Molec. gen. Genet.* **119**, 161-168 (1985).
5. Lorz, H., Baker, B. & Schell, J. *Molec. gen. Genet.* **119**, 178-182 (1985).
6. Potrykus, I., Saul, M. W., Petruska, J., Paszkowski, J. & Shillito, R. D. *Molec. gen. Genet.* **119**, 183-188 (1985).
7. Fromm, M. E., Taylor, L. P. & Walbot, V. *Nature* **319**, 791-793 (1986).
8. Vasil, I. K. in *Cell Culture and Somatic Cell Genetics* (Academic, Orlando, 1984).
9. Fujimura, T., Sakurai, M., Akagi, H., Negishi, T. & Hirose, A. *Pl. Tissue Lett.* **2**, 74-75 (1985).
10. Yamada, Y., Zhi-Qi, Y. & Ding-Tai, T. *Pl. Cell Rep.* **5**, 85-88 (1986).
11. Toriyama, Y., Hinata, K. & Sasaki, K. *Theor. appl. Genet.* (in the press).
12. De la Peña, A., Puertas, M. J. & Merino, F. *Chromosoma* **83**, 241-248 (1981).
13. Puertas, M. J., De la Peña, A., Estades, B. & Merino, F. *Chromosoma* **89**, 121-126 (1984).
14. Knop, W. *Landwirtsch. Vers. Sin.* **7**, 93-107 (1885).
15. Ozias-Akins, P. & Vasil, I. K. in *Cell Culture and Somatic Cell Genetics of Plants* (Academic, London, 1985).
16. Reiss, B., Sprengel, R., Will, H. & Schaller, H. *Gene* **30**, 217-223 (1984).
17. Schreier, P. H., Seftor, E. A., Schell, J. & Bohnert, H. J. *EMBO J.* **4**, 25-32 (1985).
18. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
19. Dellaporta, S. L., Wood, J. & Hicks, J. B. *Pl. molec. Biol. Rep.* **1**, 19-21 (1983).
20. Czernilofsky, A. P. *et al. DNA* **5**, 101-103 (1986).
21. Merino, F. thesis. Univ. Santiago de Compostela (1986).
22. McQuade, H. A. & Young-Rotter, A. M. *Am. J. Bot.* **71**, 228-238 (1984).
23. Heslop-Harrison, J. *Ann. Bot.* **30**, 221-229 (1966).

Quisqualate receptors are specifically involved in cerebellar synaptic plasticity

Masanobu Kano* and Makoto Kato

Department of Physiology, Faculty of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

Long-term modification of transmission efficacy at synapses is the cellular basis of memory and learning¹⁻⁴. A special type of synaptic plasticity in the cerebellum was postulated theoretically^{5,6}, and has since been verified^{4,7-9}. Each cerebellar Purkinje cell (PC) receives two distinct excitatory inputs, one from parallel fibres (PFs) and the other from a climbing fibre (CF). When these two types of inputs are conjunctively activated, PF-PC transmission undergoes long-term depression (LTD)^{4,7-9}. Accumulated evidence suggests that LTD plays a role in the motor learning processes of the cerebellum¹⁰⁻¹². At the molecular level, LTD appears to be caused by desensitization of receptor molecules in PC dendrites towards the PF neurotransmitter⁴, presumably L-glutamate (Glu) (ref. 12). Glu receptors are heterogeneous and can be divided into several subtypes¹³⁻¹⁵. In this study, we compared the potency of several Glu agonists in inducing LTD and found a highly selective dependency of LTD on the quisqualate(QA)-selective subtype of Glu receptors.

* Present address: Department of Physiology, Jichi Medical School, Tochigi-ken, 329-04, Japan.

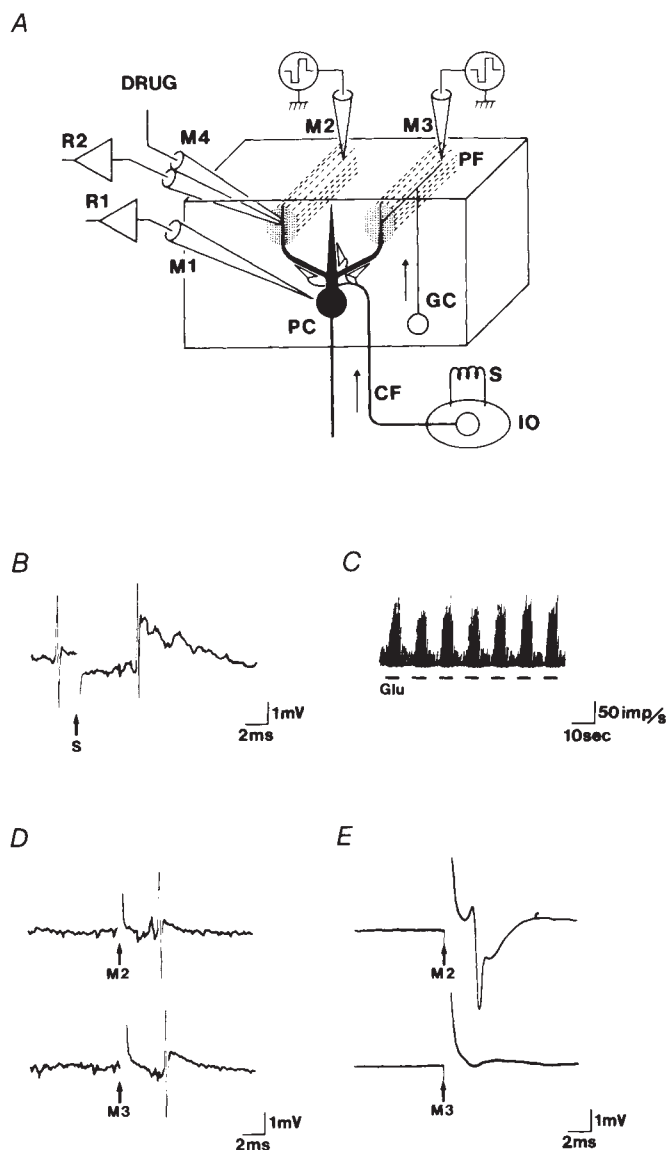


Fig. 1 Experimental system. **A**, Schematic drawing of the cerebellar cortex with recording and stimulation electrodes attached. M1, M2 and M3, glass-coated tungsten microelectrodes. The tip of M1 was exposed for 15 μm , and those of M2 and M3 for 100 μm . M4, double- or triple-barrelled glass microelectrode for application of drugs. The individual barrels of M4 contained the following solutions: L-glutamate (Glu, 0.2 M, pH 7.0), L-aspartate (Asp, 0.2 M, pH 7.0), quisqualate (QA, 0.01 M in saline, pH 6.0), kainate (KA, 0.01 M in saline, pH 7.0), *N*-methyl-D-aspartate (NMDA, 0.01 M in saline, pH 8.0), 2-amino-5-phosphonvalerate (APV, 0.2 M, pH 8.0), kynurenate (KYNA, 0.2 M, pH 8.0). R1 and R2, AC amplifiers (R1, band width, 0.08–10,000 Hz, connected to M1; R2, band width, 15–10,000 Hz, connected to the recording barrel of M4). S, bipolar platinum-iridium electrode for stimulation of climbing fibres (CFs); PC, Purkinje cell; GC, granule cell; PF, parallel fibre; CF, climbing fibre; IO, inferior olive. **B**, Extracellular recording from a PC with R1. The arrow marks the moment of IO stimulation. Note that just before the stimulus artifact, there is a spontaneous simple spike. **C**, Stripe chart record of a simple spike discharge from the same PC as **B**. Bin width, 0.5 s. Horizontal bars under the record show the periods of Glu application. **D**, Record from the same PC as **B**, but with M2 and M3 stimulation. **E**, Field potentials evoked with M2 and M3 stimulation recorded with R2. Note that marked field potentials were evoked with M2 stimulation, but not with M3 stimulation. These field potentials consist of a characteristic initial negative spike, n_1 , and a succeeding negativity, n_2 , representing PF volleys and excitation of cortical cells evoked thereby. **B** and **D** are single sweep records. **E** is the average of 8 sweeps, repeated at a rate of 0.5 s^{-1} .

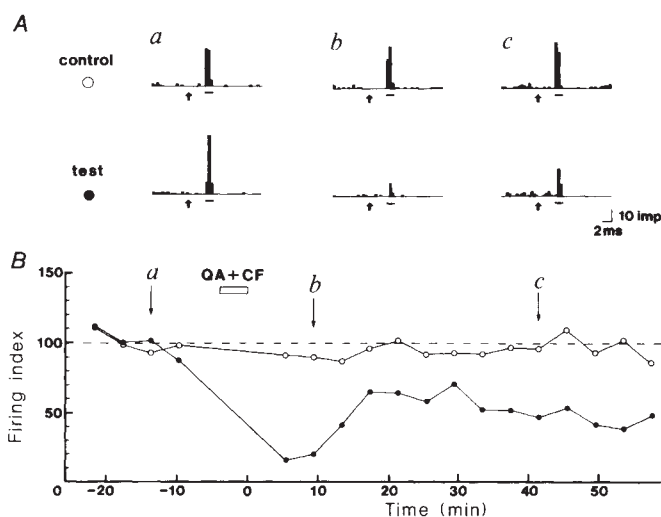


Fig. 2 Depression of PF-PC transmission induced by conjunctive iontophoretic application of QA to PC dendrites with electrical stimulation of CFs. **A**, Peristimulus time histograms of PC responses to stimulation of the control (upper) and the test (lower) PF beams, each constructed during 120 successive stimuli repeated at the rate of 0.5 Hz. Horizontal bars below histograms, stimulus-induced firing of the PC. Bin width, 0.5 ms. Arrows, moments of PF stimulation (**a**) before and (**b**) 9 min and (**c**) 40 min after conjunctive application of QA with 4 nA for 4 min with CF stimulation, as indicated in **B**. **B**, Values of the firing index for stimulation of the control ($\circ$) and the test ($\bullet$) PF beams, relative to the control values prior to the QA-CF conjunction (horizontal open bar). Times **a**, **b** and **c** are as in **A**. Firing index was calculated as the number of evoked spikes (total spike number minus the expected number of spontaneous spikes) during a 1.5 ms period (horizontal bars below histograms in **A**) divided by 120, the number of stimulation trials.

Experiments were performed on the superficial folium of the dorsal paraflocculus of decerebrate rabbits, prepared as described elsewhere^{4,7,8}. Figure 1 shows a diagram of the experimental set-up and specimen records. A glass-coated tungsten microelectrode (M1) was placed on a PC which was identified by characteristic CF responses to olivary stimulation (Fig. 1B). Two glass-coated tungsten microelectrodes (M2, M3), separated by a distance of 150–200 μm from each other, were inserted into the molecular layer at about 500 μm from the PC so as to activate the PC through separate PF beams (Fig. 1D). A drug-containing glass micropipette (M4) was placed on the dendrites of the PC being recorded, identified by responses of the PC to an iontophoretically applied Glu agonist (Fig. 1C). Then the position of M4 was adjusted so that it was on one of the PF beams, the test beam, and field potentials evoked with M2 were recorded through M4 (Fig. 1E). Stimulation of another PF beam, the control beam with M3 did not evoke any noticeable potentials (Fig. 1E). The stimulus strength for each PF beam was 4–15 μA (biphasic constant current pulses, total duration of 0.4 ms) which excited the PC in 40–100% of the stimulation trials. The efficacy of PF-PC transmission was quantified by calculating the firing index^{4,8} on peristimulus time histograms constructed from 120 successive trials of PF stimulation (Fig. 2A). The two PF beams were tested simultaneously by alternating stimulation, each at a stimulation frequency of 0.5 Hz.

When a PC exhibited stable PF responsiveness for at least 10 min, conjunctive CF stimulation with local application of a Glu agonist was performed in the following manner. While Glu, L-aspartate (Asp), QA or kainate (KA) was ejected from the drug-containing electrode by pulsative iontophoresis (pulse duration: 2–5 s), CFs were stimulated at 2 Hz. The CF stimulation was continued for 5 s and then was interrupted for 5 s. These stimulation trials continued for 4 min. Pulsative iontophoretic

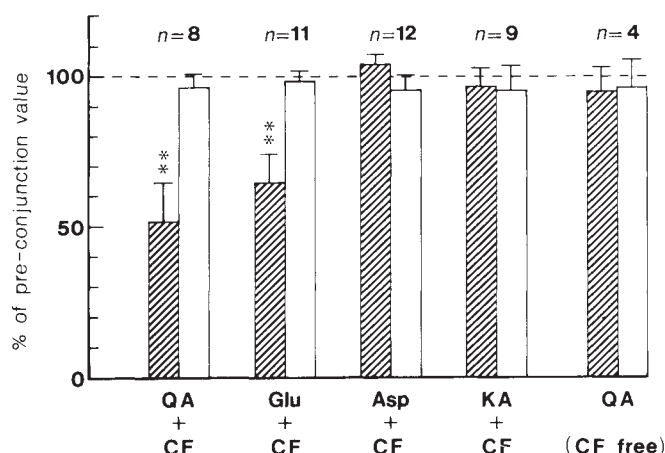


Fig. 3 Relative potency of Glu agonists. Mean percentage reduction of the values of the firing index in stimulation of the test (hatched bars) and the control (open bars) PF beams, at 20–30 min after termination of conjunction of CF stimulation and application of the Glu agonist indicated under the abscissa. Note that in the case of QA (CF free), QA alone was applied in the absence of both spontaneous and stimulus-evoked CF responses. Vertical lines, s.e.m.; *n*, number of cells tested; **, $P < 0.01$ for one-sided signed Wilcoxon test for reduction of the values of the firing index relative to its pre-conjunction value.

application of a Glu agonist was synchronized with this alternating CF stimulation, but the onset time and the duration of the drug ejection were varied in order to induce an optimal increase in the simple spike discharge from the PC during the period of CF activation (duration: 5 s for Glu and Asp, 3 s for QA, 2 or 3 s for KA). The intensity of ejection currents was chosen so that the instantaneous frequency of a simple spike discharge from the PC increased to 100–150 s⁻¹ (–5 to –30 nA for Glu, Asp and KA, –1 to –15 nA for QA; retention currents, +5 nA). Adoption of this rather complicated procedure was necessary for obtaining constant activation of the PC during conjunction. Otherwise, the PC might be depressed during continuous CF stimulation and drug application. In four experiments, the spontaneous occurrence of complex spikes was reversibly suppressed by local injection of lidocaine (4%, 2–6 µl) through a micro-syringe inserted into the region of the contralateral inferior olive (IO). QA was then applied in the absence of CF inputs with ejection current and ejection time equivalent to those used for QA–CF conjunction.

QA–CF conjunction regularly induced a depression of the responsiveness of the PC to stimulation of the test beam (Fig. 2). The depression reached its maximum within 20 min and lasted for more than 50 min. By contrast, there was no noticeable change in the PC responsiveness to stimulation of the control beam, which was located about 150 µm from the QA-ejecting electrode. Similar depression was observed with Glu–CF conjunction, but never with Asp–CF or KA–CF conjunction (Fig. 3). QA is thus the only Glu agonist which effectively depressed PF–PC transmission in a manner similar to Glu. Application of QA in the absence of spontaneous CF activity produced no effect on the PC responsiveness to stimulation of the test beam, as confirmed in four PCs (Fig. 3).

Conjunctive application of any of the four Glu agonists with CF stimulation did not affect the PC responsiveness to stimulation of the control beam. Since the drug-applying electrode was located 150–200 µm from the stimulation electrode for the control beam, the dendritic region of the PC activated by the control beam was presumably free of influences from the drug applied to the test beam region. Furthermore, QA application in the absence of spontaneous CF activity was ineffective in depressing

Table 1 Effects of APV and KYNA on responses of Purkinje cells to NMDA, QA and PF stimulation

	NMDA*	Percentage blockade QA	PF stimulation
APV; –30 nA (<i>n</i> = 8)	82.2 ± 6.3†	4.9 ± 4.7	5.8 ± 3.4
KYNA; –40 nA (<i>n</i> = 8)	—	77.4 ± 7.7†	82.3 ± 6.7†

Values are mean ± s.e.m. APV and KYNA were continuously applied (4 min with current shown) onto the dendritic region of the PC.

* NMDA evoked inhibitions on PCs.

† $P < 0.01$, one-sided signed Wilcoxon test; before versus after application of either APV or KYNA.

PF–PC transmission. Therefore, it is evident that conjunction of CF activity and QA or Glu application is indispensable for inducing a depression of PF–PC transmission.

Neither Asp nor KA induced a depression of PF–PC transmission in spite of their excitatory effects on PCs. This implies that the effect of Glu and QA in depressing PF–PC transmission does not merely result from depolarization or consequent simple spikes evoked in the PC. Both Asp-selective and KA-selective receptors appear to be present on the PC membrane^{16–19}, but these receptors are apparently not involved in the depression of PF–PC transmission.

Another Glu agonist, *N*-methyl-D-aspartate (NMDA), was not included in the test described above, because application of NMDA (ejection current: –25 to –40 nA) by pulsative iontophoresis (pulse duration: 6–10 sec) did not activate PCs, as reported previously^{16,17}. When applied to the dendritic region of a PC, NMDA usually induced an inhibition of a simple spike discharge, presumably through activation of inhibitory interneurons²⁰. In order to examine whether NMDA receptors are responsible for PF–PC transmission, experiments were performed using a selective NMDA antagonist, 2-amino-5-phosphonovalerate (APV), together with a non-selective Glu receptor antagonist, kynurenic acid (KYNA) (ref. 20). Iontophoretic application of APV (–30 nA for 4 min) significantly blocked the NMDA-induced inhibition (Table 1). However, the same dosage of APV did not antagonize the PC responses to both QA application and PF stimulation (Table 1). By contrast, iontophoretic application of KYNA (–40 nA for 4 min) significantly antagonized both PF–PC transmission and the QA-induced excitation (Table 1). These results indicate that PF–PC transmission is mediated, at least mostly, at non-NMDA receptors.

Since the depression of PF–PC transmission induced by QA- and Glu–CF conjunction is similar in time-course and magnitude to that observed after PF–CF conjunction, it follows that iontophoretically applied QA or Glu replaces electrical stimulation of PFs in induction of LTD. The present results thus strongly suggest that QA-selective Glu receptors in PC dendrites play major roles in PF–PC transmission and its LTD. The mechanism of LTD remains a matter of speculation, but Ca²⁺ influx accompanying CF responses seems to play an essential role. This is suggested by the following results. Activation of CFs evokes prolonged plateau potentials in PC dendrites²¹, which represent Ca²⁺ influx²². When these plateau potentials are suppressed by postsynaptic inhibition of PC dendrites through stellate cells, conjunctive stimulation of PFs with CFs fails to induce LTD (ref. 8). In horizontal cells of goldfish retinae, application of QA alone leads to desensitization of QA receptors²³. However, this desensitization itself is insufficient to explain the mechanism of LTD, since it recovers quickly after termination of QA application²³. In order for LTD to occur, sufficient Ca²⁺ inflow into PC dendrites is necessary. It is probable that the QA receptors on PC dendrites undergo long-lasting biochemical modification, such as phosphorylation of receptor protein, under conjoint

influences of Ca^{2+} ions from the inside and Glu or QA from the outside.

We thank Professor Masao Ito for his constant encouragement and for critical reading of the manuscript. We also thank Dr Y. Oohashi for help in statistical analyses. This study was supported by a grant (58060001) from the Japanese Ministry of Science, Education and Culture.

Received 22 July; accepted 6 November 1986.

- Bliss, R. V. P. & Lomo, T. J. *Physiol., Lond.* **232**, 331-356 (1973).
- Kandel, E. R. & Schwartz, J. H. *Science* **218**, 433-443 (1982).
- Alcon, D. L. *Science* **226**, 1037-1045 (1984).
- Ito, M., Sakurai, M. & Tongroach, P. J. *Physiol., Lond.* **324**, 113-134 (1982).
- Marr, D. J. *Physiol., Lond.* **202**, 437-470 (1969).
- Albus, J. S. *Math. Biosci.* **10**, 25-61 (1971).
- Ito, M. & Kano, M. *Neurosci. Lett.* **33**, 253-258 (1982).
- Ekerot, C.-F. & Kano, M. *Brain Res.* **342**, 357-360 (1985).
- Sakurai, M. *Neurosci. Lett. Suppl.* **22**, S26 (1985).
- Ito, M. A. *Rev. Neurosci.* **5**, 275-296 (1982).
- Watanabe, E. *Brain Res.* **297**, 169-174 (1984).
- Ito, M. *The Cerebellum and Neural Control*, 81-83 (Raven, New York, 1984).
- Watkins, J. C. & Evans, R. H. A. *Rev. Pharmac. Tox.* **21**, 165-204 (1981).
- McLennan, H. *Prog. Neurobiol.* **20**, 251-271 (1983).
- Foster, A. C. & Fagg, G. E. *Brain Res. Rev.* **7**, 103-164 (1984).
- Crepel, F., Dupont, J.-L. & Gardette, R. *Brain Res.* **279**, 311-315 (1983).
- Kimura, H., Okamoto, K. & Sakai, Y. J. *Physiol., Lond.* **365**, 103-119 (1985).
- Hemdon, R. M. & Coyle, J. T. in *Kainic Acid as a Tool in Neurobiology* (eds McGeer, G., Olney, J. W. & McGeer, P. L.) 189-200 (Raven, New York, 1978).
- Henke, H., Beaudet, A. & Cuénod, M. *Brain Res.* **219**, 95-105 (1981).
- Quinlan, J. E. & Davies, J. *Neurosci. Lett.* **60**, 39-46 (1985).
- Ekerot, C.-F. & Oscarsson, O. J. *Physiol., Lond.* **318**, 207-221 (1981).
- Llinás, R. & Sugimori, M. J. *Physiol., Lond.* **305**, 197-213 (1980).
- Ishida, A. T. & Neyton, J. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1837-1841 (1985).

Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid

P. Ross*, H. Weinhouse*, Y. Aloni*,
D. Michaeli*, P. Weinberger-Ohana*, R. Mayer*,
S. Braun*, E. de Vroom†, G. A. van der Mare†,
J. H. van Boom† & M. Benziman*‡

* Department of Biological Chemistry, Institute of Life Sciences,
The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
† Gorlaeus Laboratories, PO Box 9502, 2300 RA Leiden,
The Netherlands

Cellulose is the most abundant renewable carbon resource on earth and is an indispensable raw material for the wood, paper, and textile industries. A model system to study the mechanism of cellulose biogenesis is the bacterium *Acetobacter xylinum* which produces pure cellulose as an extracellular product. It was from this organism that *in vitro* preparations which possessed high levels of cellulose synthase activity were first obtained in both membranous and soluble forms¹⁻³. We recently demonstrated that this activity is subject to a complex multi-component regulatory system^{4,5}, in which the synthase is directly affected by an unusual cyclic nucleotide activator enzymatically formed from GTP, and indirectly by a Ca^{2+} -sensitive phosphodiesterase which degrades the activator. The cellulose synthase activator (CSA) has now been identified as bis-(3'→5')-cyclic diguanylic acid ($\text{p}5'\text{G}3'\text{p}5'\text{G}3'\text{p}$) on the basis of mass spectroscopic data, nuclear magnetic resonance analysis and comparison with chemically synthesized material. We also report here on intermediary steps in the synthesis and degradation of this novel circular dinucleotide, which have been integrated into a model for the regulation of cellulose synthesis.

A cyclic structure, consisting of two or more riboguanosine residues linked by 2'-5' or 3'-5' phosphodiester bonds, was pro-

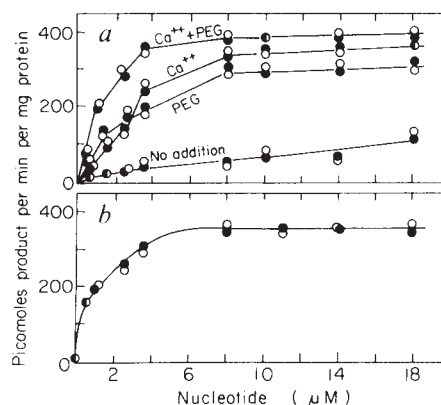


Fig. 1 Effects on activation of cellulose synthase by (●) native CSA and (○) synthetic c-di-GMP of Ca^{2+} and PEG. *a*, Synthase in washed membranes^{1,2,4}. Standard reaction mixtures contained (as shown) 8% PEG 4000, 1 mM CaCl_2 , 1 mM CaCl_2 plus PEG, or no addition. *b*, Digitonin-solubilized synthase³. Activity as presented was not affected by addition of CaCl_2 or PEG. Concentrations of nucleotide were calculated using a molar extinction coefficient of 24,700 at 252 nm, pH 7.0.

Methods. CSA was prepared enzymatically from GTP (ref. 5), modified by employing affinity purified diguanylate cyclase. Synthesis of c-di-GMP and 3'-5' cyclic triguanylic acid was by a published procedure¹² with slight modification (E. de V., G.A. van der M. and J.H. van B., in preparation). Their structures were unambiguously ascertained by ¹H- and ³¹P-NMR spectroscopy.

posed for the CSA^{4,5}. Direct support for a cyclic dimeric structure was obtained by DEAE-Sephadex chromatography⁶, which indicates that the compound contains not more than two GMP residues, and by mass spectroscopic analysis (fast atom bombardment in the negative ion mode), which indicates that its relative molecular mass (M_r) is 690, a value corresponding to that of cyclic diguanylic acid. That the CSA is indeed the 3'-5' isomer of cyclic diguanylic acid, namely bis-(3'→5')-cyclic diguanylic acid (c-di-GMP), was first indicated by its susceptibility to the 3'-5' phosphodiester-specific T1 endonuclease⁷, and finally established by chemical synthesis of the putative compound. The chemically synthesized material and the enzymatically formed activator are: (1) identical in stimulating cellulose synthase activity under a variety of conditions (Fig. 1); (2) show the same sensitivity and yield identical products when subjected to a variety of chemical and enzymatic treatments^{4,5}; (3) yield identical proton and phosphorus NMR (nuclear magnetic resonance) spectra (Fig. 2); and (4) are indistinguishable by HPLC analysis (Fig. 3b). To the best of our knowledge, our findings represent the first case of a cyclic dinucleotide of biological origin observed with a defined physiological function, namely the regulation of cellulose biosynthesis.

The enzyme responsible for the synthesis of the c-di-GMP activator, referred to here as diguanylate cyclase, has been purified 3000-fold from the supernatant of washed membranes. Results of parallel experiments, using either [γ -³²P]- or [α -³²P]GTP as substrate, indicate that in the diguanylate cyclase reaction, one mole of pyrophosphate (PPi) is released for every mole of GMP residue incorporated from GTP into c-di-GMP (Fig. 3a). In both reaction mixtures, an additional ³²P-labelled product is formed in small amounts, containing twice as much [α -³²P]GTP-derived label as that from [γ -³²P]GTP. This material, referred to as 'product C', is devoid of stimulatory activity and apparently is an intermediate in the diguanylate cyclase reaction; when subjected as sole substrate to the action of the purified enzyme, product C is completely converted to c-di-GMP (Fig. 3b). Taking together the stoichiometry of the cyclase reaction, the structure of its final product and the pattern

‡ To whom correspondence should be addressed.

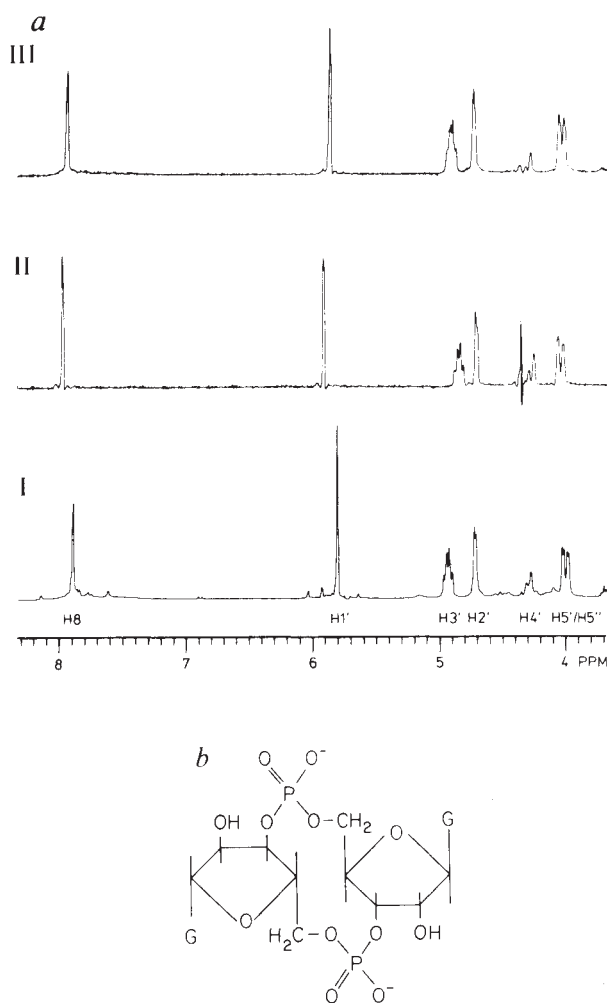


Fig. 2 *a*, NMR spectroscopy of native CSA and synthetic c-di-GMP. ¹H-NMR spectra of (I) naturally occurring CSA, (II) synthetic c-di-GMP, (III) equimolar mixture of native CSA and synthetic c-di-GMP, recorded at 50 °C (D₂O, 300 MHz); chemical shifts (δ-values) are given in p.p.m. relative to the internal standard tetramethylammonium chloride (δTMA = 3.19 p.p.m.). The assignments of the low-field ribose protons H1' to H5' are based upon 2D-COSY spectra of the synthetic and natural materials. The H1' protons of synthetic c-di-GMP and natural CSA have a coupling constant (*J*_{H1',H2'}) of 1.8 Hz. Also, ³¹P-NMR (D₂O; 80.7 MHz) spectroscopy of a mixture containing synthetic c-di-GMP and natural CSA revealed the presence of one resonance at -0.62 p.p.m. relative to the external standard aqueous (85%) H₃PO₄. *b*, Structural formula of bis-(3'→5')-cyclic diguanylic acid (c-di-GMP).

of ³²P incorporation from GTP into product C, this presumed intermediate is most likely to be the linear diguanosine-tetraphosphate pppG3'p5'G. This conclusion is borne out by the finding that synthetically prepared pppG3'p5'G is identical to product C with respect to HPLC retention and in its ability to serve as a precursor in the diguanylate cyclase reaction (Fig. 3b).

The target enzyme for c-di-GMP action is the membrane-bound cellulose synthase. Activation of this enzyme is highly specific for the dimeric structure, since neither the lower homologue, 3',5' cyclic GMP, nor the higher homologue, 3'-5' cyclic tri-guanylic acid (chemically synthesized, Fig. 1 legend) at concentrations up to 0.5 mM, stimulate the enzyme. In the entire range of c-di-GMP concentrations tested, stimulation of

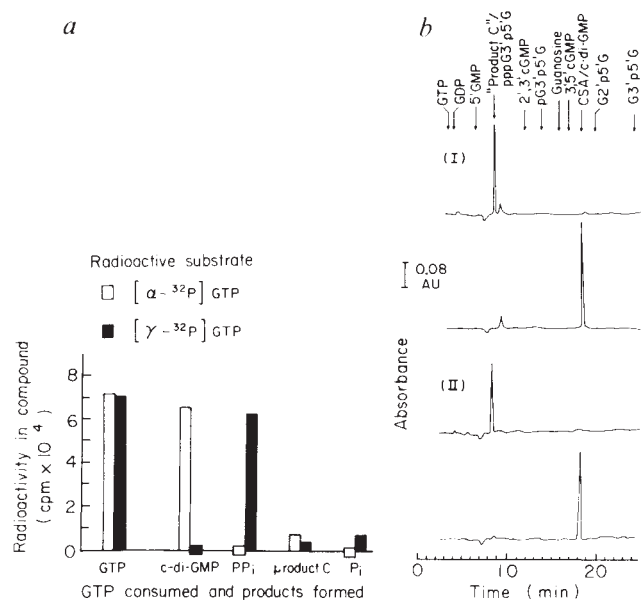


Fig. 3 Characterization of the diguanylate cyclase reaction. *a*, Stoichiometry of the reaction. The values attained for the [α-³²P]GTP reaction mixture have been multiplied by 1.2 in the diagram to correct for the higher specific radioactivity in the [γ-³²P]GTP reaction mixture. *b*, Conversion of product C and pppG3'p5'G to c-di-GMP. Diguanylate cyclase (0.1 μg protein) was incubated with either (I) 2.1 nmol of synthetic pppG3'p5'G or (II) 1.2 nmol of purified product C and 1 mM CaCl₂ in 0.05 ml TME buffer. Presented is the HPLC elution profile of the reaction mixtures at zero time (top trace in each part) and 20 min (bottom trace).

Methods. Diguanylate cyclase was purified from the supernatant of washed membranes by precipitation in 10% PEG, followed by DEAE-cellulose chromatography, employing a linear gradient of 0.05–0.3 M KCl in 60 mM Tris-HCl, pH 7.5, containing 10 mM MgCl₂ and 1 mM EDTA (TME buffer), and finally by affinity chromatography⁴, modified by including 1 mM dithiothreitol (DTT) and 10% glycerol in the GTP eluate. The GTP concentration in the enzyme-containing GTP-eluted fractions was reduced to 30 nM by vacuum dialysis. In *a*, reaction mixtures, containing 0.2 mM of either [α-³²P]GTP (100,000 c.p.m. nmol⁻¹) or [γ-³²P]GTP (120,000 c.p.m. nmol⁻¹), 3 mM sodium pyrophosphate (PPi) (to minimize degradation of the labelled PPi released), 1 mM CaCl₂, and vacuum dialysed enzyme (0.1 μg protein) in 0.05 ml TME buffer, were incubated for 15 min at 30 °C. Reaction was terminated by applying 10 μl of each reaction mixture to polyethyleneimine-cellulose plates which were developed in 5.5 M (NH₄)₂SO₄, pH 3.5, for determining c-di-GMP (*R_F* 0.03) and product C (*R_F* 0.14), and in 1.5 M KH₂PO₄, pH 3.65, for determining GTP (*R_F* 0.38); Pi (*R_F* 0.80); and PPi (*R_F* 0.66). Labelled spots, detected by autoradiography, were extracted and counted. In *b*, product C was isolated from scaled-up diguanylate cyclase reaction mixtures and purified by HPLC fractionation; pppG3'p5'G was synthesized as reported¹³. The identity and homogeneity of the product were corroborated by ¹H and ³¹P-NMR spectroscopy. HPLC was performed on a 250 × 4 mm RP-18 column (Merck) at 40 °C, detection at 252 nm, on a Merck-Hitachi liquid chromatograph. Runs were carried out in 0.15 M NaH₂PO₄, pH 5.2, at 1 ml min⁻¹, using a multi-step gradient of acetonitrile (steps are isocratic, unless otherwise stated); 1 min (0–2% linear gradient), 1–12 minutes (2%): 12–15 minutes (2–6% linear gradient), 15–20 minutes (6%), 20–25 minutes (6–0% linear gradient). Preparation of [terminal 5'-³²P]ppG3'p5'G was by 5' phosphorylation of G3'p5'G with [γ-³²P]ATP and polynucleotide kinase (Pharmacia)¹⁴, followed by purification of the product by HPLC, as above. Its structure was verified by UV spectroscopy and conversion to G3'p5'G and ³²Pi with alkaline phosphatase.

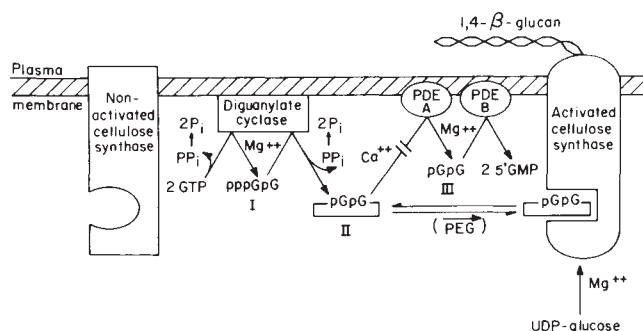


Fig. 4 Proposed model for regulation of cellulose synthesis in *A. xylinum*. Diguanylate cyclase, presumably a membrane-associated enzyme, converts two molecules of GTP into a 5' triphosphate dimer (I) which is then condensed intramolecularly to form cyclic diguanylic acid (II). The pyrophosphate released in each step is rapidly hydrolysed to inorganic phosphate. Association of cyclic diguanylic acid with the cellulose synthase activates the enzyme. (*In vitro*, this interaction is promoted by PEG.) Deactivation occurs when a membrane-bound phosphodiesterase (PDE-A) cleaves the cyclic structure of the nucleotide to yield the inactive open dimer (III). A second phosphodiesterase (PDE-B) splits this product into two molecules of 5' GMP. Calcium ions block the first of these degradative steps, thus preserving the nucleotide activator and extending its availability to the synthase. For simplicity, the synthesis of a single 1,4- β -glucan chain is depicted, although a more complex form of the synthase, polymerizing several chains simultaneously, might be the active enzyme unit in cellulose biogenesis.

synthase activity is markedly increased by the presence of either Ca^{2+} or polyethyleneglycol (PEG) (Fig. 1a). The effect of Ca^{2+} is attributed to its inhibition of the membrane-bound enzyme which degrades the activator (see below). The interaction between the activator and the synthase within the membrane is apparently reversible, since membranes exposed to c-di-GMP lose all their enhanced synthase activity upon washing, which can subsequently be restored upon readdition of the activator. Enhanced activity is not due to an increase in the affinity of the synthase for UDP-glucose, since the K_m for this substrate² (0.2 mM) is not affected by c-di-GMP. The digitonin-solubilized cellulose synthase³ is similarly stimulated by c-di-GMP (Fig. 1b). In this case, however, stimulated activity is not affected by Ca^{2+} or PEG. The absence of the Ca^{2+} effect is compatible with our finding that the enzyme degrading the activator is destroyed by digitonin treatment. Although the mechanism of cellulose synthase activation by c-di-GMP is still unclear, it is meaningful that the 1,4- β -glucan product of the cyclic nucleotide activated synthase of *A. xylinum*, recently analysed by electron microscopy and electron diffraction⁸, has been shown to consist of fibrillar cellulosic material (probably cellulose I).

The enzymatic degradation of CSA by membranous preparations of *A. xylinum* to 5' GMP as sole final product, occurs by two distinct steps⁷. The initial degradation product arising from ³²P-labelled c-di-GMP has now been isolated and identified as the linear dimer pG3'p5'G on the basis of enzymatic⁵ and sequential degradation by periodate/ β -elimination^{5,9} and by comparison to synthetically prepared (Fig. 3b) pG3'p5'G (data not shown). Thus, the first step of the pathway of CSA degradation involves hydrolytic opening of its cyclic structure. This step is strongly inhibited by Ca^{2+} (with an EDTA-free membrane preparation, this activity is inhibited 50% by 40 μM Ca^{2+}). On the other hand, the conversion of the linear dinucleotide to 5' GMP is not affected by Ca^{2+} , which may indicate the involvement of two different enzymes in the degradation of the c-di-GMP. With the activator being a cyclic nucleotide, the action of its degrading enzyme(s) may then be similar in its site of hydrolysis and presumed physiological role to that of the specific

phosphodiesterases acting on 3',5' cyclic AMP or 3',5' cyclic GMP.

A model summarizing the sequence of reactions involved in the regulatory mechanism of the cellulose synthase of *A. xylinum* is presented in Fig. 4. The pace of the polymerization process is set by the cellular levels of c-di-GMP which, in turn, depend on the concentration of cellular Ca^{2+} . The stability and unique structure of the nucleotide activator are compatible with its role as an intracellular messenger regulating cellulose synthesis *in vivo*. Although the regulation of cellulose synthesis has thus far been demonstrated only for *A. xylinum*, it is tempting to speculate that mechanisms similar to that described here, based on cyclic diguanylic acid or on related cyclic di- or oligonucleotides, may function in other organisms and other cellular processes. In this respect, it is interesting to note that the initiation reaction of RNA polymerase *in vitro*, under certain conditions, is accompanied by release of dinucleoside 3',5' tetraphosphates¹⁰ and strongly inhibited by several synthetic cyclic 3',5' dinucleotides¹¹.

We thank Professor H. Schwartz and Dr K. Eckart (Technical University of Berlin) for the fast atom bombardment mass spectroscopy analysis. This work was supported by a grant from the United States-Israel Binational Science Foundation, and by the Robert Szold Institute for Applied Science, The Hebrew University of Jerusalem.

Received 18 June; accepted 11 November 1986.

- Aloni, Y., Delmer, D. P. & Benziman, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6448-6452 (1982).
- Benziman, M., Aloni, Y. & Delmer, D. P. *J. appl. Polymer Sci.* **37**, 131-143 (1983).
- Aloni, Y., Cohen, R., Benziman, M. & Delmer, D. P. *J. biol. Chem.* **258**, 4419-4423 (1983).
- Ross, P. et al. *FEBS Lett.* **186**, 191-196 (1985).
- Ross, P. et al. *Carbohydr. Res.* **149**, 101-117 (1986).
- Tener, G. M. *Meth. Enzym.* **12A**, 394-404 (1967).
- Egami, F., Takahashi, K. & Uchida, T. *Prog. nucl. Acid Res. molec. Biol.* **3**, 59-62 (1964).
- Lin, F. C., Brown, R. M., Cooper, J. B. & Delmer, D. P. *Science* **230**, 822-825 (1985).
- Martin, E. M., Birdsall, N. J. M., Brown, R. E. & Kerr, I. M. *Eur. J. Biochem.* **95**, 295-307 (1979).
- Hinrichsen, A. I., Ortner, I. M. & Hartmann, R. *FEBS Lett.* **193**, 199-202 (1985).
- Hsu, C. J. & Dennis, D. *Nucleic Acids Res.* **10**, 5637-5647 (1982).
- Hsu, C. J., Dennis, D. & James, R. A. *Nucleosides & Nucleotides* **4**, 377-389 (1985).
- Wreessmann, C. T. J., Fiddler, A., Veeneman, G. H., van der Marel, G. A. & van Boom, J. H. *Tetrahedron Lett.* **26**, 933-936 (1985).
- Hurwitz, J. & Novogrodsky, A. *Meth. Enzymol.* **12B**, 207-212 (1968).

Initiation of translation is impaired in *E. coli* cells deficient in 4.5S RNA

David B. Bourgaize* & Maurille J. Fournier

Department of Biochemistry, University of Massachusetts, Amherst, Massachusetts 01003, USA

The 4.5S RNA of *Escherichia coli* is a small, stable RNA that is essential for cell growth but its function is not yet known. Its biosynthesis is stringently controlled, and it is processed by RNase P, a transfer RNA processing enzyme¹⁻³. To identify the biological role of the 4.5S species, we have characterized the physiological changes that occur when the bacterial cell is depleted of this RNA. We used a strain of *E. coli* in which synthesis of the 4.5S RNA can be turned off by removing an inducer of the *lac* operon, resulting in cell death⁴. We report here that an early consequence of depriving the cell of 4.5S RNA is the accumulation of translationally-defective ribosomes, which maintain their ability to elongate polypeptide chains, but can no longer participate in the initiation of protein synthesis.

Bacterial strains in which synthesis of 4.5S RNA is under the control of isopropyl- β -D-thiogalactoside (IPTG), an inducer of the *lac* operon, were constructed (see refs 4, 5 for details). A copy of the 4.5S RNA structural gene (*ffs*) was fused to an

* Present address: Alma College, Alma, Michigan 48801, USA.

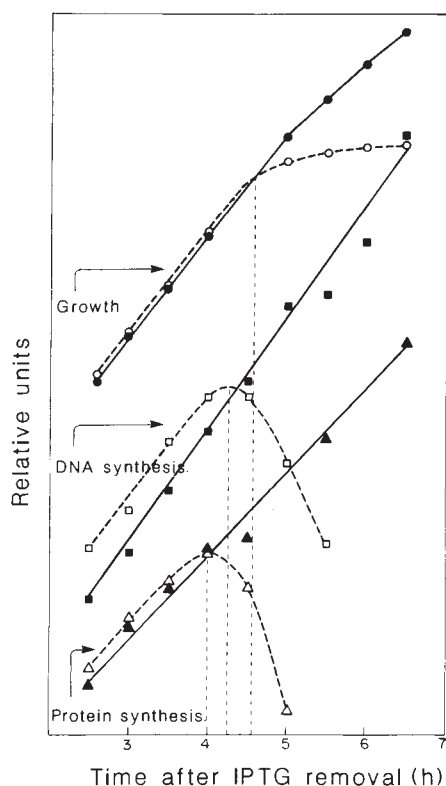


Fig. 1 Cessation of growth, DNA synthesis, and protein synthesis as 4.5S RNA is depleted. Strain FF242 was grown to saturation at 37 °C in minimal MOPS-glucose medium¹⁶ supplemented with 10 mM IPTG. After being collected by centrifugation, the cells were washed twice with medium containing no IPTG. These cells were then used to inoculate 50 ml cultures with and without IPTG. Growth was monitored at 535 nm. At the times indicated during logarithmic growth, 300 μ l of cells was removed from each culture and labelled for 5 min with either 0.1 μ Ci of [³⁵S]methionine (400 Ci mmol⁻¹), or 0.04 μ Ci of [¹⁴C]thymidine (55.7 mCi mmol⁻¹). After labelling, 3 ml of cold 10% trichloroacetic acid (TCA) was added. Samples were heated at 90 °C for 15 min, and the acid-insoluble activity collected by filtration and measured by scintillation counting. Circles, turbidity with (●) and without (○) IPTG; squares, incorporation of [¹⁴C]thymidine with (■) and without (□) IPTG; triangles, incorporation of [³⁵S]methionine (▲) with and (△) without IPTG.

IPTG-inducible *tac* promoter⁶. This *Ptac::ffs* fusion was inserted into the DNA of lambda phage cI857nin5 at the unique *Xho*I site. This phage was crossed with *λimm434*, and the resultant *λimm434 Pta::ffs* phage was packaged and introduced into the genome of *E. coli* S1224 HfrH *lacI*^Q by lysogeny. The normal chromosomal *ffs* gene was then inactivated by partial replacement of the coding sequence with a kanamycin determinant⁴, using phage P1 transduction to replace the 5' two-thirds of the *ffs* gene by a gene encoding resistance to kanamycin. The resulting strain is designated S1192. Another strain, with the same *ffs* genotype, was constructed essentially as described, but with W3110 *lacI*^Q L8 (ref. 7) as the recipient of the recombinant phage. This second strain, called FF242, permits direct comparison with strain W3110 which has been widely used for growth and metabolic studies.

Withdrawal of IPTG results in cessation of growth as 4.5S RNA is depleted (ref. 4 and Fig. 1). Gel electrophoresis of RNA extracted from cells at various times shows that the amount of 4.5S RNA decreases several fold as IPTG deprivation continues (data not shown). Incorporation of labelled methionine into protein ceases 15–20 min before either DNA synthesis or cell growth is impaired. The relative immediacy of this effect,

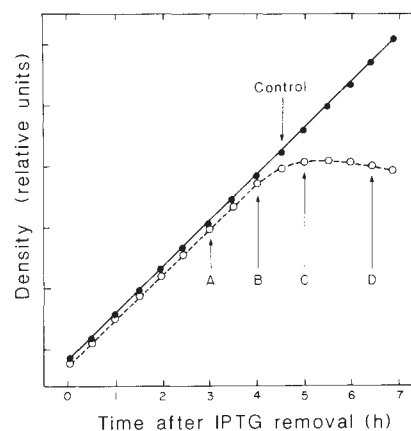


Fig. 2 Growth of cells with *lac*-controlled 4.5S RNA synthesis. *E. coli* strain S1192 was grown to saturation at 37 °C in yeast extract and tryptone (YT) medium supplemented with 0.4% lactose, instead of IPTG. Cells were transferred to YT medium without lactose for 40 min at 37 °C, then diluted into YT medium with and without 10 mM IPTG, and growth was monitored. Density markings correspond to logarithmic units. ●, Turbidity with IPTG; ○, turbidity without IPTG; A, B, C, D, times of sampling.

Methods. At the times indicated, cultures were chilled, harvested, and quick-frozen. Frozen pellets were ground with twice their weight of alumina and resuspended in 1.5 ml of 10 mM Tris-acetate, pH 8.2, 14 mM magnesium acetate, 60 mM KCl, and 1 mM dithiothreitol (DTT) per g of cells. Extracts were centrifuged for 5 min at 5,000g to remove the alumina and cell debris, and supernatants centrifuged twice for 30 min at 30,000g. To the resulting supernatants (S-30) was added 160 μ l per g of cells of 0.75 M Tris-acetate, pH 8.2, 0.021 M magnesium acetate, 7.5 mM DTT, 0.075 mM each amino acid, 6 mM ATP, 67.5 mM phosphoenolpyruvate, and 0.02 mg ml⁻¹ pyruvate kinase. Extracts were incubated for 80 min at 37 °C in the dark, then dialysed overnight against two changes of 10 mM Tris-acetate, pH 8.2, 14 mM magnesium acetate, 60 mM potassium acetate, and 1 mM DTT. Extracts were quick-frozen at -80 °C in small aliquots.

taken with the data on its regulation and processing, suggests that 4.5S RNA may be involved in protein synthesis.

The translational activities of extracts from 4.5S RNA-depleted cells were evaluated in the *in vitro* system of Zubay⁸. Extracts were prepared at various times following removal of inducer and compared with a control extract from cells maintained in the presence of IPTG. Figure 2 shows the growth rate of the cells with and without IPTG and the times at which the 4.5S RNA-depleted extracts were prepared. The translational activity of the various extracts was measured as described, using two types of RNA templates⁸. These were: poly(U), expression of which does not require the normal initiation process; and natural messenger RNA, consisting of endogenous mRNA present in the extracts and exogenous *Q β* phage RNA, for which the initiation events must occur. Figure 3a and b shows the results of testing various 4.5S RNA-depleted extracts for their ability to use each template. When poly(U) was present, all extracts were seen to have the same activity, suggesting that the capacity to elongate polypeptide chains is the same in control and 4.5S RNA-depleted extracts. However, when natural message was used, translational activity decreased as 4.5S RNA was depleted. As these extracts maintain the ability to elongate, it follows that initiation must be impaired in the depleted extracts.

We examined the structural integrity of the ribosomes in crude extracts using sucrose density gradient centrifugation. Ionic conditions were selected that do not promote association of free subunits or cause dissociation of 70S complexes⁹. Extracts dep-

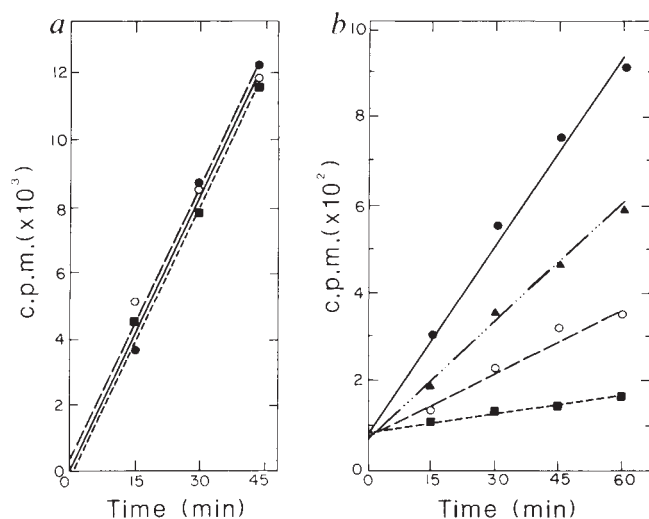


Fig. 3 *a*, Translational activity of control extract and 4.5S RNA-depleted extract using poly(U) as template, measured as incorporation of [¹⁴C]phenylalanine. ●, Control extract; ○, extract C; ■, extract D. *b*, Translational activity of control extract and 4.5S RNA-depleted extract using natural message, measured as incorporation of ³H-labelled amino acids (mix of all but asparagine, cysteine, glutamine, methionine and tryptophan). ●, Control extract; ▲, extract A; ○, extract C; ■, extract D.

Methods. Extracts were prepared as in Fig. 2, and 250 µg of the appropriate extract was added to a final volume of 200 µl, with 44 mM Tris-acetate pH 8.2, 55 mM potassium acetate, 27 mM ammonium acetate, 14.7 mM magnesium acetate, 7.4 mM calcium acetate, 1.37 mM DTT, 0.22 mM each asparagine, cysteine, glutamine, methionine and tryptophan, 2.2 mM ATP, 0.55 mM GTP, 21 mM Na₃PEP (phosphoenolpyruvate), 100 µg ml⁻¹ *E. coli* tRNA, 27 µg ml⁻¹ pyridoxine-HCl, 27 µg ml⁻¹ NADP, 27 µg ml⁻¹ flavin adenine dinucleotide (FAD), 11 µg ml⁻¹ *p*-aminobenzoic acid, 27 µg ml⁻¹ folic acid, 50 µg ml⁻¹ of either poly(U) or Qβ RNA, and either 2 µCi [¹⁴C]phenylalanine (513 mCi mmol⁻¹) or 0.2 µCi [³H] labelled amino-acid mix (250 mCi ml⁻¹ each amino acid). Extracts were added last, and 38 µl removed at various times into 1 ml of cold 5% TCA. Samples were heated at 90 °C for 15 min, and the acid-insoluble activity collected by filtration and measured by scintillation counting.

leted of 4.5S RNA contained more free subunits than control extracts, consistent with the observation that initiation is impaired in depleted extracts. Although extracts from later time points showed a greater than fourfold increase in the level of free subunits, the magnitude of this effect does not correspond to the larger decrease (90%) observed in translational activity. This suggests that a much larger portion of the ribosome pool may be unavailable for initiation.

The reduction in translational activity can be explained by a defect within the components required for initiation, or by the presence of an inhibitor that accumulates as a consequence of 4.5S RNA depletion. To distinguish between these possibilities, control extract was fractionated further into a S-150 fraction, a 1 M salt ribosomal wash fraction containing initiation factors (IF), and salt-washed 70S ribosomes. These fractions were each added separately to the least active 4.5S RNA-depleted extract, and their ability to restore activity tested (Fig. 4). The S-150 and IF fractions, observed to be functional with control 70S ribosomes (data not shown), had no effect on the activity of the depleted extract. But, the addition of an amount of washed 70S ribosomes roughly equal to that present in control extract restored activity to the control level. As the same ribosomes had little effect on the control extract, we conclude that translation in the depleted extracts is limited by a lack of initiation-

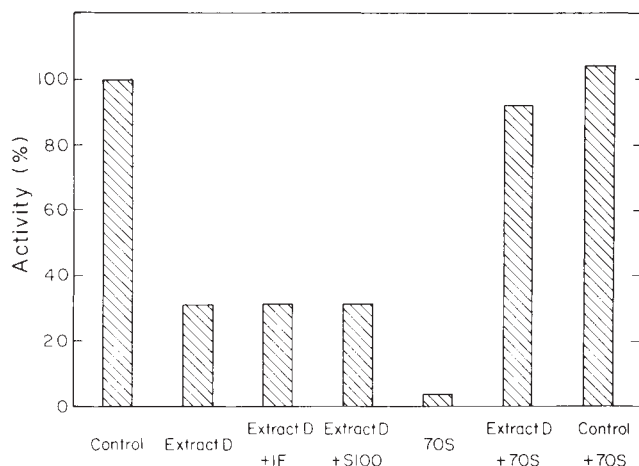


Fig. 4 Protein synthetic activity of a 4.5S RNA-depleted extract complemented with fractions of control extract.

Methods. Control extract was fractionated as follows: S-30 extract was spun at 40,000 r.p.m. for 4 h in a Beckman SW-60Ti rotor to pellet crude ribosomes. The supernatant was the S-150 fraction. The crude ribosomes were resuspended in 10 mM Tris-Cl, pH 7.6, 15 mM MgCl₂, 1.0 M NH₄Cl, 5 mM 2-mercaptoethanol, and incubated on ice for 1 h with occasional agitation. The washed ribosomes were collected by ultracentrifugation as above, and initiation factors (IF) were precipitated from the supernatant by the addition of 470 mg ml⁻¹ (NH₄)₂SO₄ (ref. 17). This precipitate was collected by (ref. 17) centrifugation and resuspended in 10 mM Tris-Cl, pH 7.6, 20 mM NH₄Cl, 5 mM MgCl₂, 5 mM 2-mercaptoethanol, and dialysed overnight against the same. Aliquots of the IF fraction were then quick-frozen and stored at -80 °C. The washed ribosomes were resuspended in 10 mM Tris-Cl, pH 7.6, 15 mM MgCl₂, 60 mM NH₄Cl, 5 mM 2-mercaptoethanol, and pelleted again as above. Aliquots were resuspended in the same buffer, quick-frozen and stored at -80 °C. Activity of control extract was defined as 100%. Reactions were carried out as in Fig. 3, using Qβ mRNA, except that no time points were taken; after 30 min the entire reaction mix was added to cold 5% TCA. The following additions were made where indicated: 80 µg S-150, 80 µg IF, 80 µg washed 70S ribosomes.

competent ribosomes. Apparently, the ribosomes become defective as 4.5S RNA is depleted.

The possibility that the functional defect in the ribosomes of 4.5S RNA-depleted cells could be correlated with a structural alteration was assessed by sucrose density gradient analysis. No difference in sedimentation was observed for either intact ribosomes or individual subunits from depleted extracts when compared with control ribosomal particles (data not shown). Thus, if a structural alteration is the basis of the effect, it does not result in grossly modified subunits, although more subtle alterations are not ruled out.

No increase in activity was observed when purified 4.5S RNA was added to depleted extracts, even in amounts 100-fold greater than that typically found in control or wild-type extracts. The 4.5S RNA normally accounts for 1–2% of the total small RNA in *E. coli*¹⁰, and is present in extracts at about 1 µM. These results suggest that the 4.5S species does not act as a classic soluble factor in the initiation of protein synthesis. Nor does the 4.5S RNA seem to serve as an essential structural component of the ribosome, because the wild-type level of 4.5S RNA is approximately one tenth that of the 5S ribosomal RNA, and virtually all 5S RNA is ribosome associated¹¹. Earlier experiments have shown that the 4.5S RNA is not associated with crude ribosomes^{12,13}.

Because the 4.5S RNA cannot be shown to restore activity to depleted extracts by itself, we cannot conclude that the observed effect is a direct result of 4.5S RNA depletion. However, less direct evidence argues strongly for the participation of this

molecule in the protein synthetic process. 4.5S RNA is under stringent control, as are the ribosomal and transfer RNAs (tRNAs), and also like tRNAs, is processed by RNase P. It has a long lifetime, as do other RNAs involved in protein synthesis¹³. More direct is the observation that the earliest event following 4.5S RNA depletion is the appearance of initiation-incompetent ribosomes. Three hours after the removal of IPTG, which is roughly three generations, initiation activity is 65% that of control. The longevity of the molecule (longer than that of 5S RNA, ref. 13) suggests that this number of cell divisions must occur before dilution and turnover cause the cellular level to drop below that required for proper function. Also, there is no detectable alteration in the protein makeup of cells, as judged by two dimensional electrophoresis, until after initiation becomes impaired (D.B.B., T. A. Phillips, F. C. Neidhardt & M.J.F., in preparation). Patterns of labelled proteins remain constant until initiation activity drops below the 65% level. Decreases below this point are accompanied by an increase in expression of several proteins. There is no apparent accumulation of incomplete peptides. Taken together, these correlations support the view that impairment of translational initiation occurs as a direct consequence of the loss of 4.5S RNA.

The data suggest that the 4.5S RNA participates in creating or maintaining the viability of a ribosomal subunit. That ribosomal particles appear normal, as judged by the somewhat insensitive sedimentation assay, suggests that a late stage of assembly or maturation is affected. It is unlikely that ribosomal RNA maturation is affected by a lack of 4.5S RNA; if assembled, ribosomes containing unprocessed RNA would show altered sedimentation. Evidence that the 4.5S RNA is not a structural component comes from the fact that active ribosomal particles can be assembled *in vitro* from individual RNAs and proteins without 4.5S RNA being present^{14,15}. Of course, this finding does not preclude the possibility that 4.5S RNA might have a catalytic role in assembly *in vivo*. Thus, 4.5S RNA may not be required for initial assembly or maturation, but for the development or maintenance of a functional form as subunits recycle into each initiation phase. Interestingly, the 4.5S RNA contains exposed sequences similar to the Shine-Dalgarno sequence of mRNA. Two such overlapping domains occur in the apex region of the predicted hairpin-like secondary structure and this region is particularly susceptible to nuclease (ref. 10 and B. Sneath, M.J.F. & J. N. Vournakis, in preparation). Perhaps the 4.5S RNA binds to the 3' end of the 16S RNA and assists in mRNA binding or prevents the degradation of the 3' end of the 16S RNA. In the absence of 4.5S RNA, the 3' end of the 16S RNA could be degraded, resulting in an inability to participate in initiation.

Experiments are under way to determine which subunit contains the defect responsible for translational inactivation, which event in the initiation process is affected by this defect, and whether conditions can be found in which 4.5S RNA will restore activity to depleted extracts. Results from these experiments will allow the models discussed above to be tested.

We thank Stanley Brown for providing *E. coli* strain Sf192, advice on construction of strain FF242, and for stimulating discussions. This work was supported by NIH USPHS Grant GM19351.

14. Held, W. A. & Nomura, M. *Biochemistry* **12**, 3273-3276 (1973).
15. Spillman, S., Dohme, F. & Nierhaus, K. H. *J. molec. Biol.* **115**, 513-535 (1977).
16. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. *J. Bact.* **119**, 736-747 (1974).
17. Chambliss, G. H., Henkin, T. M. & Leventhal, J. M. *Meth. Enzym.* **101**, 598-605 (1983).

Erratum

A cell surface molecule distributed in a dorsoventral gradient in the perinatal rat retina

Martha Constantine-Paton, Andrew S. Blum, Rosalia Mendez-Otero & Colin J. Barnstable

Nature **324**, 459-462 (1986).

In this letter the wrong Fig. 4 was used. This is the correct Fig. 4 and its legend, which is as published.

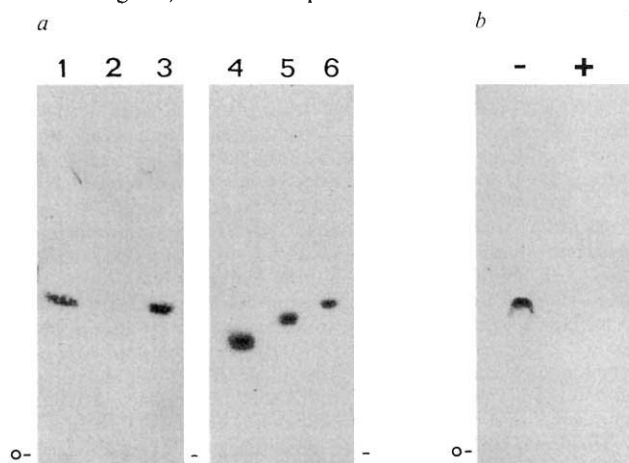


Fig. 4 Biochemical characterization of the JONES antigen using TLC/immunoblot analysis. *a*, The JONES antigen partitions into the aqueous phase of a Folch extract and is detected as a single band on immunoblots of lipids resolved by TLC. This band migrates to a position between that of G_{M2} and G_{M1} ganglioside standards run at the same time. (1) Whole lipid extract; (2) organic phase; (3) aqueous phase; (4) G_{D1a} ; (5) G_{M1} ; (6) G_{M2} . 1-3, Immunostaining of the nitrocellulose blot of a TLC plate; 4-6, resorcinol staining of purified standards run at the same time. *b*, The epitope recognized by JONES antibody contains sialic acid as it is removed by treatment with neuraminidase (+) but not buffer (-).

Methods. P4 rat retinas were homogenized in chloroform/methanol (2:1) to give a crude lipid extract³⁶. This was partitioned into organic and aqueous phases by 0.1 M KCl⁹. Gangliosides and other highly polar lipids (including some phospholipids) partition into the aqueous phase, whereas most neutral glycolipids and some less polar gangliosides partition into the organic phase. The sialic acid content of the fractions was determined by the resorcinol method³⁷. Samples containing 0.3-0.4 μ g sialic acid were applied to a precoated HPTLC plate (Merck, silica gel 60, 0.2 mm thick) and developed with chloroform/methanol/0.25% aqueous $CaCl_2$ (60:40:10). Chromatograms were dried, wetted with isopropanol/water (2:1) and blotted onto nitrocellulose³⁸. The nitrocellulose blot was blocked in 5% normal goat serum and incubated with JONES antibody for 4 h. After washing, the blot was incubated in peroxidase-conjugated goat anti-mouse IgG (H + L) for 2-4 h and washed. Bound antibody was visualized by incubation with 3,3' diaminobenzidine. Ganglioside standards were run on parallel lanes as controls and were visualized by resorcinol staining³⁷. For neuraminidase digestion the aqueous phase retinal lipids were dried under N_2 and resuspended in 0.1 M sodium acetate, pH 5.5, with or without neuraminidase (Sigma type X from *Clostridium perfringens*) at 8 mU per μ g sialic acid. Incubations were carried out for 15 h at 37 °C. Digested lipids were extracted into chloroform: methanol, 2:1, dried under N_2 and resuspended in a small volume of chloroform: methanol, 2:1. Samples were run and analysed by TLC/immunoblot as described above.

Received 18 August; accepted 22 October 1986.

1. Ikemura, T. & Dahlberg, J. *J. biol. Chem.* **248**, 5033-5041 (1973).
2. Ikemura, T., Shimura, Y., Sakara, H. & Ozeki, H. *J. molec. Biol.* **96**, 69-86 (1975).
3. Bothwell, A. L. M., Garber, R. L. & Altman, S. *J. biol. Chem.* **251**, 7709-7716 (1976).
4. Brown, S. & Fournier, M. J. *J. molec. Biol.* **178**, 533-550 (1984).
5. Brown, S. & Fournier, M. J. (manuscript in preparation).
6. Brosius, J. & Holy, A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 6929-6933 (1984).
7. Brent, R. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4204-4208 (1981).
8. Chen, H.-Z. & Zubay, G. *Meth. Enzym.* **101**, 674-690 (1983).
9. Baranov, V. I., Belitsina, N. V. & Spirin, A. S. *Meth. Enzym.* **59**, 382-397 (1979).
10. Hsu, L. M., Zagorski, J. & Fournier, M. J. *J. molec. Biol.* **178**, 509-531 (1984).
11. Lindahl, L. *J. molec. Biol.* **92**, 15-37 (1975).
12. Griffin, B. E. *J. biol. Chem.* **250**, 5426-5437 (1975).
13. Lee, S. Y., Bailey, S. C. & Apirion, D. *J. Bact.* **133**, 1015-1023 (1978).